GENERAL LIBRARY
NOV 1921
UNIV. OF MICH.

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN
WALTER S. MONROE

VOL. IV OCTOBER, 1921 No. 3

CONTENTS

MEASURING THE PROGRESS OF PUPILS BY MEANS OF STANDARDIZED TESTS ..... Samuel S. Brooks 161
VARIATION OF MARKING SYSTEMS AS DIAGNOSED BY OBJECTIVE TESTS ..... R. H. Jordan 173
THE RATE OF PROGRESS IN TEACHER PREPARATION ..... W. Randolph Burgess 180
SCHOOL VARIATION IN GENERAL INTELLIGENCE ..... Warren W. Coxe 187
SOME FURTHER STUDIES OF GIFTED CHILDREN ..... Elizabeth Cleveland 193
MOTIVATED DRILL WORK IN THIRD-GRADE ARITHMETIC AND SILENT READING ..... J. H. Hoover 200
SOME ELEMENTARY STATISTICAL CONSIDERATIONS IN EDUCATIONAL MEASUREMENTS ..... J. Crosby Chapman 212
EDITORIALS 221
REVIEWS AND ABSTRACTS 229
NEWS ITEMS AND COMMUNICATIONS 233
NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH 239

Entered as second-class matter at the Post Office at Menasha, Wisconsin, under the act of March 3, 1879.

Journal of Educational Research

Copyright 1921, by
Public School Publishing Co.

Official Organ of the National Association of Directors of Educational Research

Editor-in-Chief: **B. R. Buckingham**
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

Associate Editors

E. J. ASHBAUGH, *Ex-Officio, Secretary of Research Association, University of Iowa, Iowa City.*
W. W. CHARTERS,
Carnegie Inst. Tech., Pittsburgh.
S. A. COURTIS,
City Schools, Detroit, Mich.
GEORGE D. STRAYER,
Columbia University, New York City.
LEWIS M. TERMAN,
Leland Stanford Junior University, Stanford University, California.
GUY M. WHIPPLE,
Director, Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Michigan.

Contributing Editors

VIRGIL E. DICKSON,
Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.
CHARLES FORDYCE,
Dean, Teachers' College, and Director of Bureau of Educational Measurements, University of Nebr.
M. E. HAGGERTY,
Dean of the College of Education, University of Minnesota.
V. A. C. HENMON,
Director, School of Education, University of Wisconsin.
ARTHUR W. KALLOM,
Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.
F. J. KELLY,
Dean, School of Education and Director of Bureau of School Service, University of Kansas.
ROBT. H. LANE,
Assistant Superintendent of Schools, Los Angeles, Cal.
WILLIAM A. McCALL,
Assistant Professor of Education, Teachers' College, Columbia University.
GEORGE W. MELCHER,
Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.
EUGENE A. NIFENECKER,
Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City.
JOSEPH P. O'HERN,
Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.
LEON. O. SMITH,
Assistant Superintendent (in charge of Educational Research), Public Schools, Omaha, Nebr.
W. W. THEISEN,
Director, Division of Reference and Research, Public Schools, Cleveland, Ohio.
CLIFFORD WOODY,
Professor of Education, University of Washington.

PUBLISHERS' NOTICE

THE JOURNAL OF EDUCATIONAL RESEARCH is edited by Doctor B. R. Buckingham and a representative staff of associate and contributing editors. It is published by the Public School Publishing Company and is the official organ of the National Association of Directors of Educational Research. In the fullest sense it belongs to the subscribers and to those who contribute to its pages.

The agreement to publish a magazine for the Bureau of Educational Research of the University of Illinois has been annulled in accordance with its terms. In fulfillment of our obligations, assumed in that agreement to our subscribers, we now furnish this Journal published under the same editor and with a representative editorial staff.

But in establishing a Journal independent of any single institution, we wish to acknowledge a debt which we feel educational progress owes to the University of Illinois for its early recognition of the efforts which so many workers in educational research were ready to make in support of a journal in this field. The name of the magazine may be changed with the next volume, beginning with the January number, in order to give expression to its representative character. If our subscribers have a preference for the name, the publishers and the editor will be very glad to receive an expression of their desires.

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV OCTOBER, 1921 NUMBER 3

MEASURING THE PROGRESS OF PUPILS BY MEANS OF STANDARDIZED TESTS[1]

SAMUEL S. BROOKS

District Superintendent, Winchester, New Hampshire

From time out of mind the estimate of a pupil's progress in his school work has been left to the more or less excellent judgment of his teacher, a judgment often warped by personal prejudice due to his behavior in school, his personal appearance, or his father's standing in the community. The fact that the teacher gave tests and ranked the child on the quality of his reactions to them does not necessitate a modification of the above statement. For those tests were based solely on what she *judged* the child ought to know concerning the various school subjects as a result of her particular line of instruction. She had no way of *knowing* definitely what a child of his age and grade really ought to know in order to be as well informed as other children of his age and grade in other schools. Even the grading of the papers, after they were corrected, was mostly a matter of judgment, as has been previously shown.

Some of the more unthinking teachers took the testing and grading very seriously, marked the papers very carefully on a percentage basis, and then "passed" the pupil or "flunked" him according to whether his mark was 70 or only 69. Others, realizing more or less vaguely the injustice of such a procedure, graded the pupils' work as excellent, good, fair, poor, or very poor, which they could probably have done just as accurately without giving any tests for grading purposes at all.

But there is no longer any valid excuse for such haphazard methods of measuring the results of teaching in elementary schools. The standardized tests and scales furnish us with definite

[1] This is the fifth article by Superintendent Brooks on the general topic "Putting Standardized Tests to Practical Use in Rural Schools."

norms of achievement by means of which we can compare any child's work with the median or average for his age or grade and decide justly as to whether or not he is making normal progress.

One of my purposes in using tests has been to measure the progress of pupils in their studies. Thus far we have given the tests four times in all the schools of the district. They have been given at intervals of several months so as to permit progress between tests to show plainly in the graphs. All the data from these several tests are graphically recorded and on file. The records are very interesting and highly satisfactory so far as proof of the efficiency of this method of measurement of progress is concerned although, of course, they do not always show satisfactory progress on the part of the pupils.

As heretofore stated, our plan is to give standardized tests in as many of the elementary-school subjects as possible to all the pupils in the district three times a year. They were given first in September, 1919 for grading purposes and to get a starting point from which to measure progress. In February, 1920 the tests were given again in order to find out how the pupils were progressing and particularly to discover along what lines, if any, unsatisfactory progress was being made, so that the teachers might see where increased effort or change of method was needed. In June, 1920 they were given a third time for promotion purposes.

The scores of the individual pupils in these tests were recorded, on 4×6 cards, in the form of graphs. Each time a new test was given a new graph was drawn on each pupil's card in a different color, so that at the end of June I had, for each pupil in the district above the first grade, a graph card which showed at a glance his standing in all the subjects tested for three different periods in the school year. Each teacher had duplicate cards for the pupils of her particular school.

Since my last article was written, I have devised and had printed a 5×8 graph card which is considerably more convenient than the makeshift in use last year. The graphs reproduced in this article are shown on the new form. This new card contains not only the names of the tests but also the standard scores for each of them. Directly below the name of each test is a vertical line upon which the standard scores for that test are printed at the intersections of the vertical line with the horizontal grade

lines. For instance, the sixth-grade standard score for comprehension in Monroe's Silent Reading Test is 21. Accordingly, this number is printed at the intersection of the sixth-grade line with the vertical line below "Comprehension" and under the name of that test. The fourth-grade standard score for Woody's Division Scale is 5. The figure 5 is therefore printed at the intersection of the fourth-grade line with the vertical line directly beneath "D" under "Arithmetic-Woody." Since in the Ayres Spelling Scale and in the Hahn-Lackey Geography Scale the standard scores for any particular grade vary with the column used for testing, no scores could be printed for these tests. So, merely for convenience, the Roman numerals marking the grade lines were repeated at their intersections with the verticals for these two tests. The lowest score on any test line shows the lowest grade in which that test is given. For example, Woody's Division Scale is not given below the third grade. Hence, the lowest score for this test (3) is on the third-grade line. Similarly, Starch's History Test is not given below the sixth grade.

Figures 1, 2, and 3 are copies of the graph records of three different children for the school year 1919–1920. All three were taught by the same teacher throughout the year. The graphs are given with explanations and comments for the purpose of showing a method of recording results so as to indicate at a glance how the pupils were progressing in their school work and when they were ready for promotion.

Figure 1 shows the record of an eleven-year-old girl of about average mentality. Her mental age (M.A.) was 11 years, 7 months, and her intelligence quotient (I.Q.) was 105. Hence she is a little above the average in intelligence. Her graph, resulting from the September tests and represented by the dotted line in Figure 1, falls about equally above and below the fourth-grade line. That is, she averaged about fourth-grade (end of year) ability in the tested subjects at the beginning of the school year. Hence she was placed in the class that was beginning fifth-grade work, namely, the fifth grade according to the plan discussed in my third article. The dashed line represents the scores of the the same child from the February tests and the solid line those from the June tests. The progress of the child in her studies is shown by the steady movement of the graph from below upward.

Only two subjects show little or no increase and those will be explained a little farther on.

Let us consider separately the progress made by this pupil in each subject beginning with reading. I depend mainly on Monroe's test for measuring silent reading ability. It is well standardized, perfectly objective, eliminates the memory factor, and is, to my mind, best fitted for my particular scheme. The pupil's score for rate of silent reading in September was 80. The first point, therefore, on the September curve was plotted at the intersection of the fourth-grade line with the test line, 80 being the fourth-grade standard score as shown on the card. Her score for comprehension was 17, which is halfway between the standard scores for the fourth and fifth grades. Hence, the second point on the September graph is located halfway between the fourth- and fifth-grade lines. Now note the space between the two points just located and the corresponding points on the dashed curve. This space shows the progress made by the pupil in silent reading during the first half of the school year in relation to normal annual progress represented by the distance between the two grade lines. The advance in rate of reading is particularly marked, covering as it does the space of a grade and a quarter in a half year. The advance made in comprehension is normal; that is, a half grade of progress in a half year of work.

As shown by the corresponding points on the solid-line curve, the pupil's rate of reading increased very little during the last half of the year, while progress in ability to comprehend what was read continued to be normal. The rapid increase in rate of reading was undoubtedly due to the special emphasis placed on efficient silent reading drill which was inaugurated in the Fall term and continued throughout the year. There had never before been any such drill in any of the schools. For the year, this child's progress was a grade and a half or 50 percent above normal in rate of reading and just a grade, or normal, in comprehension.

On the addition line, note the drop of the February curve below the one for September. There might be several reasons for this, the most plausible being that the child was tired or not feeling well at the time that particular test was given in February. This surmise is supported by the fact that she "came back" strong in the June tests and showed a half grade of progress for the year in addition ability.

Little progress was shown in subtraction ability; none at all for the first half of the year. But you will note that she was already up to fifth grade in both subtraction and addition at the beginning of the year. When a child's graph shows that he is well up to or above grade in any subject, the time and effort of that child is diverted to some subject in which he is below grade. One of the chief values of the tests is their diagnostic value in showing up the weak and strong places in the work of pupils or classes so that the teacher and superintendent may know where their efforts should be concentrated in order to bring about results as nearly uniform as possible. The tendency of the graphs to flatten out and more nearly approximate a straight line toward the end of the year is the direct result of this policy of placing the emphasis where it is most needed, the places where it is most needed being indicated by the earlier graphs. The ideal curve would, of course, be a straight line, denoting ability exactly equal to the grade norms in all subjects. And an ideal year's record for a fifth-grade pupil would be three straight lines the

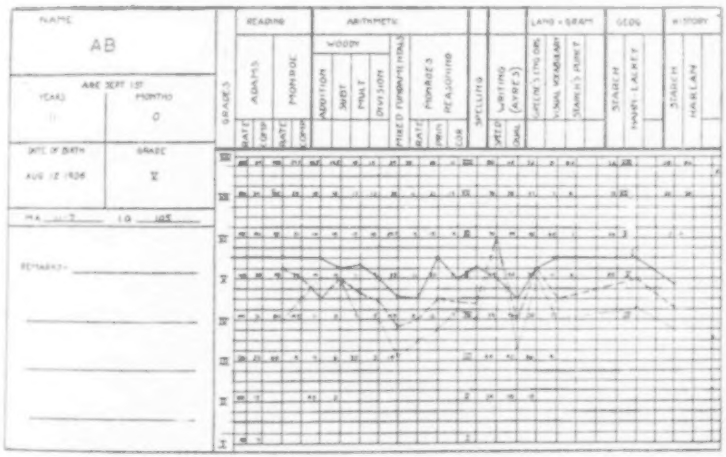


FIGURE 1. RECORD OF AN ELEVEN-YEAR-OLD GIRL OF AVERAGE ABILITY. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

first coincident with the fourth-grade line on the card, the third coincident with the fifth-grade line, and the second midway between and parallel to the others. Such a record would denote absolutely even and normal progress for the year.

One of the tests given in the fall was the Cleveland Survey Test in the fundamental operations of arithmetic—a test which is excellent for purposes of diagnosis. This test showed this particular pupil to be especially weak in the multiplication and division of fractions, decimals, and denominate numbers. Special corrective drill on these phases of arithmetic was responsible for the splendid progress shown on the multiplication and division lines.

Note the very low score made in the mixed fundamentals test in September, the excellent progress made during the year, and the fact that in spite of such progress the pupil failed to come up to grade at the end of the year. It is noteworthy that only eight pupils in the whole district have so far succeeded in getting as high grades in this test as they averaged in the four fundamental operations, although the test is made up of a mixture of the identical examples used in the addition, subtraction, multiplication, and division tests. Most of them fall below from half a grade to a whole grade. A study of Figures 2, 3, and 4 reveals the same facts concerning the results from this test. Although good progress is made in every case, the pupil or class persistently grades lower in this test than in the others on fundamentals of arithmetic. To my mind this indicates that the standard scores for this test are too high.

Continuing the examination of Figure 1, we find Monroe's Reasoning Test in Arithmetic to be the next in order. This test is scored for three things: rate of solving problems, solutions correct in principle, and correct answers. Good progress is shown for the year in all three although the pupil fails to reach the grade standard for speed in solving problems.

In spelling ability the pupil accomplished 25 percent more than a normal year's progress, with nearly four times as great progress made in the last half of the year as in the first half. And here is a chance for some more interesting comparisons of the graphs on the different cards. Figure 2 shows no progress in spelling in the first half; Figure 3 shows the same; while Figure 4, which is the record of a whole fifth grade, shows considerably more progress in

the last half than in the first. The midyear tests revealed the fact that spelling work in general was progressing unsatisfactorily. As remedial measures, oral spelling drill together with Buckwalter's *Comprehensive Speller* were thrown into the discard. Ayres' Spelling Scale, supplemented by individual spelling lists made up of troublesome words from the pupils' own written vocabularies, was made the basis of the spelling course. A little booklet containing graded lists of 1600 "Common Blunder Words" was also used in most of the schools. Spelling lessons were shortened; new words were presented by a more psychological method; and the recitation consisted of a written lesson wherein the pupils use the words of the day's lesson in sentences or in a short composition. The efficacy of these changes in subject matter and method is strikingly evidenced by the greatly increased progress during the last half of the year.

Next comes handwriting. This pupil's scores in writing are typical of the general conditions revealed by the tests as discussed in my last article; speed scores up to or much above grade and quality scores very low. Although this pupil showed considerable progress for the year, she failed to reach the grade standard in quality of handwriting. But she did better than most of the pupils in this respect. Note that, throughout the year, her speed decreased while her quality increased. In the past, speed had been attained at the expense of quality. Now quality has been gained at the sacrifice of speed, and yet speed has not been reduced below the grade standard. Figure 2 also shows the fact that quality improved at the expense of speed. In most other cases, however, speed increased at approximately the same rate as quality so that the pupils were about as far behind in writing at the end of the year as they were at the beginning. All four of the records presented in this article show an improvement in handwriting for the year considerably above the average for the district. In general the improvement in writing ability was small. The reasons for the conditions found to exist at the beginning of the year and the general lack of progress during the year were fully discussed in the preceding article.

As for language and grammar, so far as the author is aware, no satisfactory general test or scale has been standardized. One of our greatest needs at present in carrying out a complete testing program in the elementary schools is a general language and

grammar test somewhat on the same plan as the Hahn-Lackey Geography Scale. Starch's Punctuation Scale is good for measuring ability in that particular. Charters Diagnostic Language and Grammar Tests are excellent as far as they go, and they cover pretty well the common errors in the use of the English language. But no standards were available for them last year, so that they did not fit into a scheme which required tests that have been fairly well standardized.[2] Hence we could do little in testing language and grammar ability last year. The two tests used, namely, Greene's English Organization Test and Thorndike's Visual Vocabulary Test might perhaps more properly be placed under the head of reading. The English Organization Test proved rather unsatisfactory. It does not seem to measure any definite ability. Its chief value seems to be in indicating, to some extent, a pupil's general intelligence or general reasoning ability, if there is such a thing, and even in this I have not found it to agree very well with the results of regular intelligence tests.

The vocabulary test, however, has proved very valuable, especially in interpreting silent reading scores. There is a high degree of correlation between the scores in the vocabulary test and those of comprehension in silent reading if the scores of children much below normal are thrown out. When a normal child fails in comprehension of silent reading, an examination of his vocabulary scores will often show a serious lack of word knowledge, which can be remedied by a definite plan of vocabulary building as explained in a former article. To such a policy is due the excellent progress shown by the pupil represented in Figure 1 as regards vocabulary knowledge. This progress is shown by the curves to be from fourth-grade ability in September to halfway between fifth- and sixth-grade ability in June. Notice that this is also the highest point reached in the silent reading scores. This test likewise measures the efficiency of whatever method of vocabulary building may be adopted.

Highly satisfactory in amount and uniformity was the progress in geography and history, as shown in Figures 1, 2, and 4, although for some reason the history scores persistently lagged behind those in geography.

[2] Standard scores for these tests are now available and we are using them as a part of our testing program.

As before mentioned, Figure 1 is the record of a pupil a little above the average in intelligence and her record shows on the average, a little more than a normal year of progress which is as it should be. Furthermore, her progress was in the direction of a more uniform ability in all subjects. The June curve is 35 per-cent shorter than the September curve as shown in Figure 5 (a), thus approaching much nearer the ideal curve. This fact exemplifies the value of corrective measures based on diagnosis by standardized tests.

These records are also used for promotion purposes. When a child's graph has moved upward over a space approximately equal to the distance between two grade lines he is ready to be promoted to the next grade. As before stated, the pupil whose record is shown in Figure 1 was started on fifth-grade work at the beginning of the school year. Her graph has moved upward, as shown by the solid-line curve, until it averages better than fifth grade. This shows that she had attained fifth-grade end-of-the-year standards in June and was ready for promotion to the sixth grade and to begin work in that grade the following September.

Figure 2 shows the record of a very bright eleven-year-old girl with a mental age of fifteen years and an I. Q. of 135. Although her graph showed an average of sixth-grade ability at the beginning of the year, it was considered wisest, because of her youth and various changes in the course of study, to have her take the regular sixth-grade work for that year and to prepare herself for double promotion by taking part of the seventh-grade work. Her chart shows a progress of from half a grade in rate of silent reading and spelling to two and a half grades in multiplication. In the June tests, as shown by the solid-line curve, she averaged halfway between seventh- and eighth-grade standards and was promoted to the eighth grade. Whatever of seventh-grade work she did not take along with the sixth-grade work, she will take up in the eighth grade, thus losing nothing of subject matter and gaining a whole year's time. Figure 5 (b) shows the relative lengths of this pupil's September and June curves when straightened out. The June curve is about three-fourths as long as the September curve.

Figure 3 gives the record of a very dull boy with a chronological age of 13 years, a mental age of 9 years 10 months, and an I. Q. of 76. Note the great irregularity of the September curve

and the general lack of progress throughout the year. Note that in many instances the scores of later tests fall below those of previous ones, and that the reading scores are much lower than the vocabulary scores indicating that poor reading may be due to lack of native ability and not to lack of word knowledge. This boy fell so far short of reaching fourth-grade standards in the June tests that he was not promoted to the fifth grade. He was already two years retarded. Question:—Did we do right in retarding this child another year? Problem:—What to do with cases of this kind in rural schools where special classes are out of the question, where manual trade schools are beyond the reach of the pupils, when promotion means placing the pupil wholly out of his depth, and when retardation means discouragement. This boy will probably never get beyond the fourth or fifth grade except through mistaken charity. Would it not be well to have some provision whereby such hopelessly retarded children could be permitted to leave school and engage in some useful and profitable work under the guidance of parents or other responsible persons, at least until society becomes sufficiently civilized to make provision at public expense for the proper training of such individuals? They would at least be saved from forming habits of failure and idleness which so many such children acquire during years of forced attendance at school after they have reached the limits of their mental capacities in acquiring knowledge from books. Figure 5 (c) shows the relative lengths of this pupil's September and June curves. It should be remembered that all three of these pupils were taught by the same teacher in the same way.

Figure 4 is the record of a fifth grade containing nine pupils. It shows that the entire grade has made normal progress or better in nearly every test. As usual, however, the class is weak in quality of handwriting. It is also slightly below grade in arithmetical reasoning, in mixed fundamentals, in spelling, and in geography. On the other hand, the class is considerably above standard in reading, in the fundamentals of arithmetic, in speed of writing, and in language and grammar. On the whole it shows that both teachers and pupils have done excellent work throughout the year. Relative lengths of September and June curves are shown in Figure 5 (d). The June curve is about 20 percent shorter than the September curve.

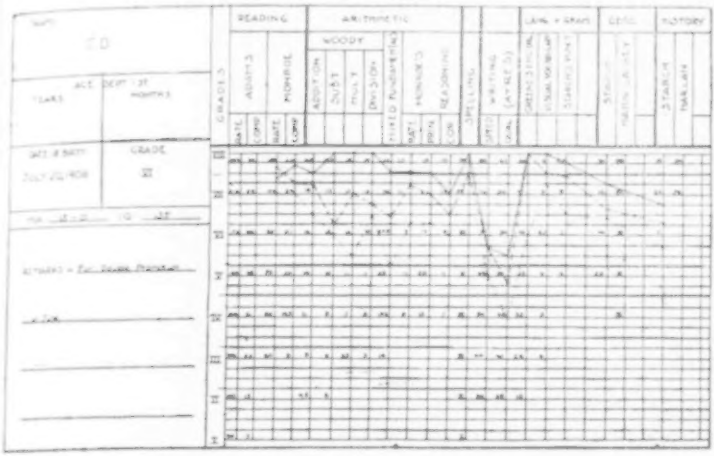


FIGURE 2. RECORD OF A BRIGHT ELEVEN-YEAR-OLD GIRL. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

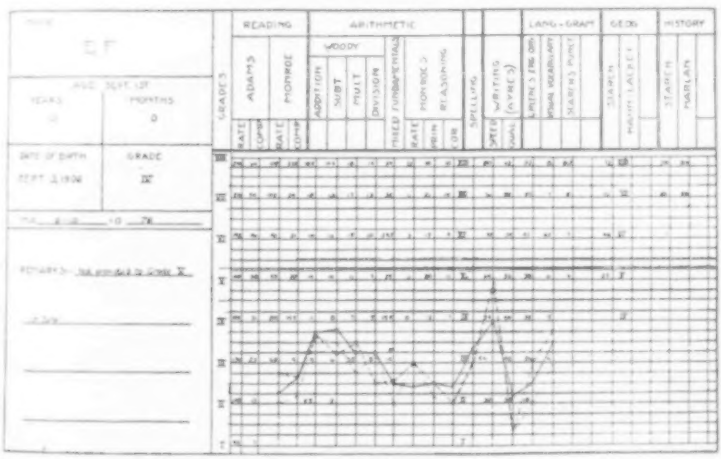


FIGURE 3. RECORD OF A VERY DULL TWELVE-YEAR-OLD BOY. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

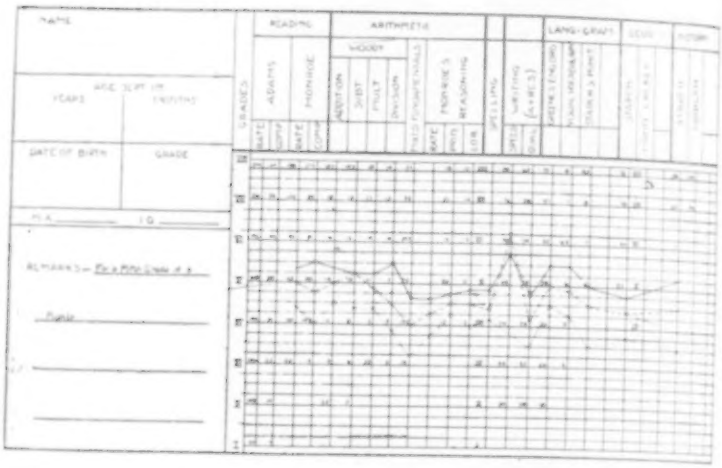


FIGURE 4. RECORD FOR A FIFTH GRADE CONSISTING OF NINE PUPILS. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

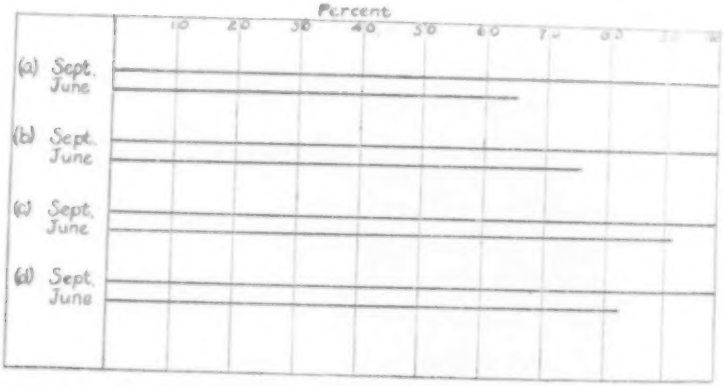


FIGURE 5. A COMPARISON OF THE LENGTHS OF SEPTEMBER AND JUNE CURVES AS SHOWN IN FIGURES 1 TO 4

VARIATION OF MARKING SYSTEMS AS DIAGNOSED BY OBJECTIVE TESTS

RIVERDA HARDING JORDAN
Dartmouth College

In recent years attention has frequently been called to the value of objective tests in diagnosing and evaluating various factors in connection with school marking, which otherwise must remain purely matters of conjecture, and the remedy for which must be applied from a purely subjective judgment. In the course of a study worked out recently, an effective concrete illustration developed which demonstrates this value very definitely. By way of bringing this use of the objective scale more especially under the observation of the school administrators, a brief description of the application made of the objective standards is here presented.

In the course of a comparison between nationality of pupils and their progress in school,[1] made in the cities of Minneapolis and St. Paul, it became necessary to collect the school marks for one semester (half a year) of the pupils in the sixth, seventh, and eighth grades. These school marks were taken from the teachers' classroom registers for ten schools in Minneapolis, involving records of 2,076 pupils. The marks were for the entire semester in each subject of instruction, and an average of all subjects for each pupil was made. In the ten schools the subjects of instruction were the same for all the pupils of any given grade, although, of course, the usual differences in the presentation of subject matter, and, indeed, in the content of subject matter offered, were found. After the average semester mark of each pupil in all his subjects was worked out, the marks were compared, and translated, where necessary, into terms of a scale of ten—i.e., a perfect mark in all subjects was 10.0, the next step in the descending scale being 9.9, and so on down to zero. Table I was then worked

[1] Jordan, R. H., Nationality and school progress, Bloomington, Illinois: Public School Publishing Co., 1921.

TABLE I. DISTRIBUTION OF MEDIAN MARKS FOR TEN MINNEAPOLIS SCHOOLS, SHOWING MEDIANS FOR BOYS AND FOR GIRLS OF EACH GRADE AND FOR EACH SCHOOL

School Number	VIIIA		VIIIB		VIIA		VIIB		VIA		VIB		All		Total
	B	G	B	G	B	G	B	G	B	G	B	G	B	G	
1[a]	7.4	8.2	7.4	8.6	6.5	8.2	6.5	8.4	6.6	6.8	7.0	7.7	6.8	8.0	7.4
2	7.8	6.8	5.5	8.4	6.5	6.6	6.4	7.0	6.7	7.8	7.8	7.3	6.8	7.3	7.1
3, 10[b]	6.0	6.7	4.7	6.7	6.6	7.6	6.3	7.0	6.2	6.5	6.3	7.7	6.0	7.0	6.5
4	5.2	6.5	5.7	6.1	5.9	6.7	6.4	6.7	5.6	6.7	7.0	7.7	6.0	6.7	6.4
5	6.2	6.7	5.1	6.4	5.2	6.4	4.8	6.5	5.5	7.6	7.5	6.4	5.7	6.7	6.2
6	5.4	7.2	4.3	6.2	5.6	7.0	4.7	6.3	4.7	7.5	3.7	7.2	4.7	6.9	5.8
7	5.7	6.7	5.4	6.7	4.8	5.8	4.7	5.3	4.9	5.3	5.2	5.3	5.1	5.9	5.5
8	5.2	6.4	4.3	6.1	4.5	5.5	5.4	6.9	5.0	3.5	5.2	5.7	4.9	5.7	5.3
9	4.5	5.7	3.0	4.5	3.0	3.1	4.4	6.7	3.7	5.3	4.8	5.2	3.9	5.1	4.5
Range	4.5 to 7.8	5.7 to 8.2	3.0 to 7.4	4.5 to 8.6	3.0 to 6.6	3.1 to 8.2	4.4 to 6.5	5.3 to 8.4	3.7 to 6.7	3.5 to 7.8	3.7 to 7.8	5.2 to 7.7	3.9 to 6.8	5.1 to 8.0	4.5 to 7.4

a Schools 3 and 10 are combined under one principal.
b Numbers given the schools arbitrarily to conceal their identity.

out on the basis of the medians of the boys and girls in each one of the school grades studied in each of the ten schools.

A study of the medians in the various schools, grade by grade, as well as a comparison of the school medians in the final column, brings to light a startling lack of uniformity in the marking systems employed in various buildings, although within any one building no great deviations will be noted. This condition gives rise to the feeling that in a school system of any size the old adage, "As the superintendent, so the school," must be changed to read, "As the principal, so the school."

The range of marking in the ten schools will be better understood by a study of Table II and Figure 1. It will be noted here that some of the marks run very low and some very high. As a

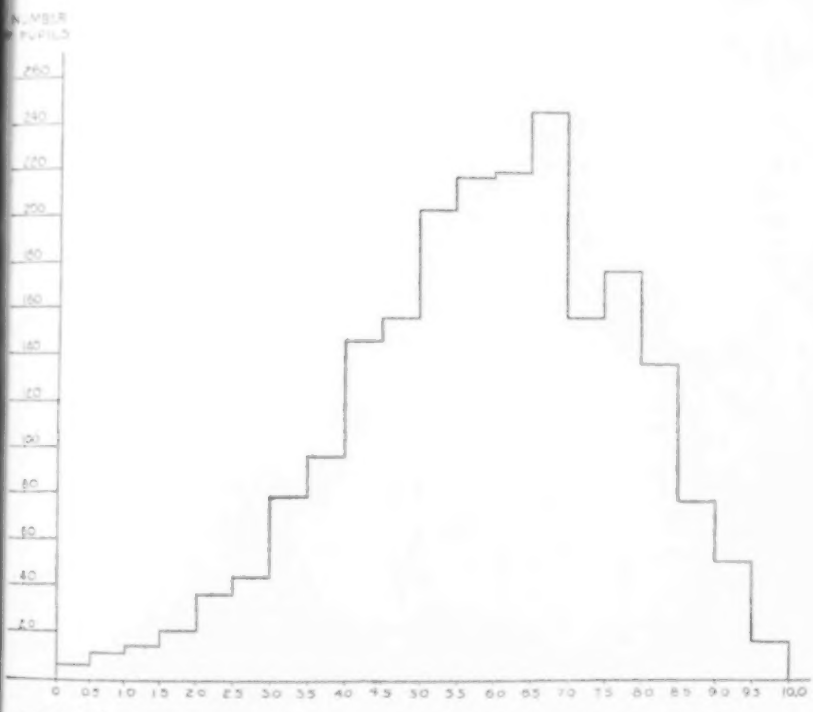


FIGURE 1. DISTRIBUTION OF SCHOOL MARKS. 10 SCHOOLS OF MINNEAPOLIS. 2,076 CASES

matter of fact, School No. 9 gives its highest ranking pupils in every grade (except the girls of VIA and VIB), a mark lower than the median pupil in the corresponding grade of School No. 1. The lower range of the scale, then, would be almost eliminated if School No. 9 were not included in the comparisons.

TABLE II. DISTRIBUTION OF AVERAGE MARKS MADE BY 2,076 PUPILS IN GRADES VIB TO VIIIA INCLUSIVE IN TEN MINNEAPOLIS SCHOOLS, IN ALL SUBJECTS

Average Mark	Boys	Girls	Total Cases
9.6–10.0	0	16	16
9.1– 9.5	7	44	51
8.6– 9.0	22	54	76
8.1– 8.5	43	91	134
7.6– 8.0	57	120	177
7.1– 7.5	61	95	156
6.6– 7.0	96	147	243
6.1– 6.5	99	116	215
5.6– 6.0	110	104	214
5.1– 5.5	107	92	199
4.6– 5.0	101	52	153
4.1– 4.5	101	45	146
3.6– 4.0	68	28	96
3.1– 3.5	55	24	79
2.6– 3.0	32	15	47
2.1– 2.5	25	10	35
1.6– 2.0	14	5	19
1.1– 1.5	10	0	10
0.6– 1.0	4	2	6
0.4– 0.5	4	0	4
Total	1,016	1,060	2,076

In conversation with the principals of such widely varying schools as School No. 1 and School No. 8 or School No. 9, an attempt was made to determine the basis which the principal had in mind in directing the marking system of the teachers under his

charge. The principal of School No. 1 felt that the high grades given by his teachers were to be explained on a basis of the high intelligence shown by the pupils of the school, and an attempt at justification of this position was made by calling attention to the fact that the pupils came from homes of intelligence and a certain degree of wealth. The principal of School No. 9 felt that the "standard" of his school was best maintained by a very strict marking system, and was very proud of the fact that pupils, who had been given higher marks in their own buildings, were not able, when transferred from other parts of the city, to meet the requirements of his own school, and frequently had to be demoted. He felt that a low marking system was an evidence of a high standard for his building. Such responses were typical of all the schools, and it became very evident that the problem of the superintendent of bringing about a uniform marking system was almost impossible unless the principals could be convinced by some purely objective method that their theories of marking were in error.

Some months later it became possible for the writer to evaluate by objective methods the achievement of a portion of the pupils already studied, by means of certain tests designed to measure specific abilities of the children. By this time, the eighth-grade children had been promoted to the high school, so that the tests were given only to the pupils who had originally been studied in the sixth and seventh grades. The tests given were two Trabue Completion tests, two vocabulary tests, two number completion tests, two geometrical forms tests, two substitution tests, a memory span test and an opposites test. All pupils were eliminated from these tests whose marks had not been considered in the original investigation. The averages of the schools is given in Table III in terms of the raw scores for each test. For puposes of comparison these tests were worked out for schools 1, 8 and 9 only, in order to determine the relative standing of these most widely divergent marking systems.

This comparison shows a clear superiority of the pupils in School No. 8 over School No. 1, and certainly demonstrates that the contention of the principal of School No. 1, that his pupils were of higher intellectual grade than those of most of the other

schools of the city, is entirely without foundation. School No. 9 is clearly below the grade of School No. 1, but not nearly to such

TABLE III. AVERAGE SCORES FOR EACH TEST IN SCHOOL 1, 8, 9

School	Opp.	Trab.	Vocab.	Subst.	Mem. Sp.	No. Comp.	Geom. Forms
1	49.7	12.6	56.9	74.9	15.2	12.3	7.1
8	57.3	13.3	58.1	79.9	16.1	12.5	8.2
9	45.9	11.5	54.4	71.0	16.4	10.1	5.5

an extent as is indicated by the wide disparity between the marks given the pupils in the two schools. The objective tests have here solved what otherwise would be an extremely difficult situation for a school superintendent to handle. It is shown very clearly that there is no justification for an assumption on the part of any one of these principals that his subjective measurement based upon long experience is a safe or proper basis for marking his pupils.

The principal of School No. 1 will at once discover his error in assuming that his pupils are superior to the personnel of the other schools of the city. The principal of School No. 9 will realize that his low marking is due in part to a lower standard of intelligence, and therefore that it does not imply necessarily a higher standard for his school, and he will see further that an extremely low standard of marking is not justified in comparison with the other schools of the city. The principal of School No. 8 likewise will realize either that his marking scale has been unfortunate or that he is not getting the degree of attainment from his pupils which he may reasonably expect.

As a result of these tests, school principals will find it incumbent to come to some agreement among themselves as to marking standards. Not only is the result of these tests of values with reference to the general conditions as brought out by medians or by averages, but the extreme range of marking shown by the total distribution of the 2,076 pupils will be very much restricted. The curve of distribution will much more nearly approach the

normal curve when the results of the objective tests are studied and intelligently applied. The ease of application of such tests as those used, or similar intelligence tests, makes it a very simple matter for any superintendent of schools to diagnose the exact situation, and hence to evaluate properly differences in marking which exist within his school system. It is to be hoped that such diagnostic methods will be used very commonly in the immediate future.

THE RATE OF PROGRESS IN TEACHER PREPARATION[1]

W. Randolph Burgess
Formerly Russell Sage Foundation

Table I shows the improvement which has taken place in the education of teachers in service in ten states since the year 1910. The index numbers[2] used to show the standing of each state are derived from state school reports which give the number of college graduates and the number of normal-school graduates on the teaching staff. The index numbers show the number of years of college or normal-school training the average teacher has had, counting only years of study leading to graduation. In Massachusetts in 1910, for example, the average teacher had one and two-thirds years of college or normal-school training. By 1918 this record had improved; so that if the completed normal-school and college courses were evenly distributed among all the teachers there would be more than two years for each teacher.

In Table I a number of missing years have been filled in by interpolation and in the case of West Virginia the index number for 1910 was estimated. The figures have also been carefully edited for inconsistencies. The states are listed in the order of their rank in 1918.

An average index number for the ten states is given at the foot of the table. This was computed from the index numbers, giving each state equal weight, a method which gives a more stable figure than a weighted average which is largely influenced by fluctuations in one or two states with large numbers of teachers. For the years 1919 and 1920 the average is computed by the method of relatives; that is the percentage increase over the 1918 level is computed in each state where the data are available, and the average of these percents is used to determine the average index figure.

The average shows a continuous improvement amounting in the aggregate to a 42 percent increase over the 1910 figures. In the light of a recent nation-wide teacher shortage it is notable that the

[1] The second of two articles on the education of teachers in service in the United States.

[2] Burgess, W. Randolph. "The education of teachers in fourteen states," *Journal of Educational Research*, 3:161–72, March, 1921.

TABLE I. INDEX NUMBERS FOR EDUCATION OF TEACHERS IN SERVICE IN TEN STATES. 1910 TO 1920

State	Years of Completed College or Normal-School Training per Teacher										
	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920
Massachusetts	1.65	1.65	1.68	1.68	1.75	1.82	1.90	1.98	2.05	2.05	2.05
Rhode Island	1.78	1.81	1.85	1.86	1.87	1.91	1.94	1.97	2.00	.....	.....
New Jersey	1.51	1.54	1.62	1.75	1.75	1.83	1.90	1.94	1.95	1.96	1.99
New Hampshire	1.00	1.09	1.17	1.23	1.30	1.36	1.43	1.51	1.59	1.65	1.71
Minnesota	0.96	1.00	1.07	1.07	1.13	1.16	1.19	1.22	1.25	1.26	1.28
Montana	0.99	1.18	1.22	1.19	1.15	1.19	1.06	1.19	1.18	1.17	1.17
Illinois	0.37	0.38	0.49	0.77	0.81	0.82	0.82	0.98	1.05	1.02	.....
Virginia	0.63	0.61	0.67	0.73	0.66	0.73	0.79	0.77	0.82	0.85	.....
Kansas	0.45	0.50	0.55	0.58	0.62	0.63	0.65	0.72	0.79	.....	.....
West Virginia	0.33	0.34	0.47	0.53	0.54	0.55	0.60	0.70	0.76	.....	.....
Average	0.97	1.01	1.08	1.14	1.16	1.20	1.23	1.30	1.34	1.35[a]	1.37
Percent increase since 1910	.....	4	11	18	20	24	27	34	39	40	42

[a] Not the averages of figures listed for 1919 and 1920, but computed from percentage increases shown by available figures in past two years.

records do not show any general falling off in the past two years. In Montana and Illinois the 1919 figures are slightly under those for 1918. In the other states there is either no loss or a slight gain; so that the average shows a slight gain. The rapid progress of previous years is interrupted, but there is no evident loss. While this showing is exceedingly encouraging it should be borne in mind that some of the effects of high prices and low teachers' salaries have not yet been felt. Decreased enrollment in normal school lessens the supply of trained teachers, not this year but in two or three years. It remains to be seen whether salary increases and better working conditions will come into play rapidly enough to offset the unfavorable influences. Apparently the situation is now better, in these states at least, than has commonly been thought.

Measuring Rates of Change

The figures of Table I are of significance, not alone for the accomplishments which they show, but also for the direction and rate of movement which they record for each state. From the

data of Table I, coefficients of regression[3] have been computed to show for each state the annual rate of progress in teacher preparation. These coefficients are shown in Table II. Illinois showed the most rapid progress during the ten-year period with an annual advance of more than seven hundredths of a year of training for each teacher. The rate is due to low figures in the first few years rather than to consistently rapid progress. It should be added that the internal evidence for the accuracy of the Illinois figures is less convincing than in the case of any other state. New Hampshire follows closely on the heels of Illinois with a remarkably steady increase in the education of its teaching staff. Montana shows the least progress, although its ranking in teacher preparation has been at or near the top of the western states for which we have figures. The state of Minnesota shows exactly the same regression coefficient as the average for the ten states.

TABLE II. COEFFICIENTS OF REGRESSION SHOWING ANNUAL RATE OF PROGRESS IN THE EDUCATION OF THE TEACHING FORCE IN TEN STATES, 1910 TO 1920

State	Annual Increase in Years of College or Normal School Training per Teacher
Illinois	0.073
New Hampshire	0.070
New Jersey	0.051
Massachusetts	0.050
West Virginia	0.048
Minnesota	0.041
Kansas	0.037
Virginia	0.025
Rhode Island	0.025
Montana	0.006
Ten states	0.041

[3] For the method employed see article by Leonard P. Ayres in *Journal of Educational Research*, May 1920, and discussion in *Trends of School Costs* by W. Randolph Burgess, Russell Sage Foundation, 1920.

Looking Forward to 1950

The coefficients of regression of Table II make possible a speculation as to the education of teachers in coming years. They furnish a method of determining where the present rate of progress will carry us. Table III shows what the education of the teachers of ten states will be in the year 1950 if the rate of progress of the past eleven years is consistently maintained.

TABLE III. INDEX NUMBERS FOR TEACHER PREPARATION IN 1950 ON THE BASIS OF THE RATE OF PROGRESS FROM 1910 TO 1920

State	Years of Completed College or Normal-School Training per Teacher
New Hampshire...	3.82
New Jersey.....	3.59
Massachusetts.....	3.58
Illinois.....	3.33
Rhode Island.....	2.78
Minnesota.....	2.58
West Virginia.....	2.26
Kansas.....	1.94
Virginia.....	1.61
Montana.....	1.36
Ten states.....	2.60

The figures are in terms of index numbers showing the years of completed college or normal-school training per teacher. The record of the highest state, New Hampshire, is very close to that goal so often mentioned, four years of college or normal-school training. Montana, the low state, shows a figure only a few points higher than her index in 1920. The present rate of progress in three states in the lower half of the table will not bring them in thirty years to a point where their average teacher will have two years of college or normal-school training.

Relative Standing at Three Periods

The relative ranking of the ten states at the three periods 1910, 1920, and 1950, according to the size of their index numbers

is shown in Table IV. Rhode Island drops from first place in 1910 to second in 1920 and fifth in 1950. Montana shows as great a decline in relative position, from fifth to tenth place. New Hampshire and Illinois show the greatest increases in rank.

TABLE IV. RANK OF TEN STATES IN TEACHER PREPARATION, 1910, 1920, AND 1950 (COMPUTED)

State	Rank		
	1910	1920	1950
Rhode Island	1	2	5
Massachusetts	2	1	3
New Jersey	3	3	2
New Hampshire	4	4	1
Montana	5	6	10
Minnesota	6	5	6
Virginia	7	8	9
Kansas	8	9	8
Illinois	9	7	4
West Virginia	10	10	7

Rapid progress is not confined to any one section of the country nor is slow progress characteristic of some single section. West and East divide the honors of the two states making most rapid progress and each furnishes a representative for lowest place. There is no tendency for urban states to make more rapid progress than rural. One of the two states with the lowest rates of gain is Rhode Island, the most thickly populated state in the Union, and the other is Montana, largely rural.

Another factor which might be expected to influence the rate of gain is the rate of increase in the population. Montana is a striking case in point. From 1910 to 1920 the population of Montana increased 45.6 per cent, an increase nearly twice as large as that in any other of the ten states. It is evident that with such an influx of new population the problem of finding trained teachers for rapidly growing schools is far more difficult than that of finding them in a more static population. It would seem reasonable to account for Montana's low rate of increase in teacher preparation by her large increase in population. When we exam-

ine this relationship in the other states, however, we find no clear-cut tendency. If the figures for Montana be omitted there is practically no correlation either positive or negative between rates of increase in population and rates of increase in teacher preparation. If there is a tendency it is toward more rapid improvement in teacher preparation in the active states having large population increases.

Conditions of Progress

The causes operating towards the improvement of the education of the teaching force can probably be determined accurately only by a more detailed analysis of circumstances in particular states. Clearly enough the first step in any such analysis is a careful compilation of the facts as to teacher preparation from year to year and a thorough interpretation of the figures once they are compiled. At present data on the educational preparation of teachers are collected in less than one-third of the states of the Union and in only a few of these states are the figures subjected to careful scrutiny that their true import may be discovered. Notable examples of careful collection and use of such figures are found in recent school reports from New Jersey, Massachusetts, and Montana.

It is a truism in education that as is the teacher so is the school. The development of modern educational statistics in the past decade has taught us the value of child accounting, and we have learned that conditions affecting the progress of children are most rapidly improved by the simple method of checking up the facts each year with regard to the ages, the grades, and the progress of the children and printing them in our annual reports. We need to learn the same lesson with regard to the teachers and to develop methods of teacher inventories which shall annually gather and publish the facts that will tell us where the teachers come from, how well they are prepared, how long they serve, and what they are paid.

Summary

1. From 1910 to 1920 there is an increase of 42 percent in the average index number of teacher preparation for ten states.

2. In the years 1919 and 1920 little progress was made in teacher preparation, but on the other hand there was no substantial loss in any state for which we have figures.

3. When computed as a coefficient of regression the annual increase in years of college or normal-school training per teacher for ten states is 0.041.

4. If present rates of progress are continued until 1950, New Hampshire will lead the states for which we have figures with average teacher preparation nearly equivalent to four years of college or normal-school training.

5. There are no obvious explanations for the very diverse rates of progress among states. The situation clearly calls for careful collection and analysis of statistics in each state.

SCHOOL VARIATION IN GENERAL INTELLIGENCE

WARREN W. COXE

Formerly Vocation Bureau, Cincinnati, Ohio

A course of study which shall be of universal application throughout a school system has certain obvious advantages, among which are the following: 1) inasmuch as all pupils of the same grade study the same material at the same time, transfers can be effected easily and without loss of time to the pupil; 2) there can be a greater uniformity of textbooks; 3) courses of study can be worked out in greater detail; 4) many supervisory problems are simplified; and 5) the cost of education is somewhat less. All these advantages are on the side of less work for the teacher and less cost for the community.

We should not forget, however, that student bodies differ greatly in general intelligence and therefore in their ability to pursue a given course of study. Were we to judge the efficiency of teaching by the degree of success with which pupils pass standard educational tests, we might be unfair unless we made allowance for such a factor as the general intelligence of the pupils. It is, therefore, very important that a school superintendent should know the composition of the student bodies in the several schools in terms of general intelligence.

This article presents data regarding the general intelligence of 24 sixth grades in 24 elementary schools in Cincinnati. The examination used was the Otis Group Intelligence Scale. It was given near the end of the year, was administered throughout by two people trained in giving such tests, and was scored by the same two people. The conditions of giving and scoring were thus uniform in all the schools. The results ought therefore to be strictly comparable.

The immediate reason for giving the tests was the selection of candidates for a six-year classical high school. The tentative basis of selection was a minimum intelligence quotient of 110.

In Table I is presented the distribution of scores for the 24 schools. The wide variation of scores in any one school is at once noticeable. Schools are not alike in this respect, for some manifest much more variation than do others. Reduction of this

TABLE I. DISTRIBUTION OF OTIS SCORES OF TWENTY-FOUR SIXTH GRADES

Scores	Schools																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
180–184																								
175–179																					1			
170–174																								
165–169																								
160–164	1																							
155–159																								
150–154	2					1				1		1					2			1				
145–149	2			2			1		1											1				
140–144	4			2	1						3	3				2	1		1	1	1			
135–139	1	2	1				1			4	4	2					1				1			
130–134	7	1	1			1			3	2		3	1		1		1				4		1	
125–129	8	1	1		3	1				1	4	10	1			2	2			3	3			
120–124	4	6	1	3	1	1	1		1	3	1	8	1	1	1	5	1		1	3	7	2	1	
115–119	9	4	3	3	1	4	4		1	1	7	3	1		2	1	2	2	1	3	3	1	2	2
110–114	6	4	6	3	1	3	10		1	7	3	1	1	1		5	8	1	1	6	5		1	1
105–109	9	9	2	11	4	6	7	1	3	6	4	10	1	1	3	3	6		1	5	6		2	4
100–104	6	3	7	3	3	6	5	1	2	3	4	9	4		2	8	7	1	6	5	10	2	3	3
95–99	10	9	7	5	5	4	2	2	6	10	7	7	3	2	5	5	8	3	3	12	11		3	3
90–94	8	6	7	2	7	3	3	1	6	6	1	8	2	1	4	9	8		4	8	7	2	2	7
85–89	13	8	7	2	4	9	7	3	4	9	5	5	4	1	3	10	4	1	4	10	5	2	9	4
80–84	7	8	8	4	2	7	6	3	4	13	8	9	1		4	9	11	3	3	9	10	6	1	2
75–79	7	6	8	2	4	4	3	2	4	10	3	7	1	6	4	8	5	7	6	6	11	6	7	2
70–74	4	11	8	2	1	5	9	6	1	8	5	4	5	2	2	5	7	8	2	5	7	4	3	
65–69	4	7	5	1	2	6	5	3	6	6	3	5	2		2	8	8	2	2	6	10	5	3	4

Scores	Schools																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
60–64	8	2	4	4	1	1	1	4	2	5	....	2	2	3	2	7	6	7	4	1	4	3	6	1
55–59	3	3	2	3	2	....	1	5	1	4	1	1	2	....	....	....	5	4	3	4	1	6	6	2
50–54	3	1	....	....	3	1	3	7	1	....	....	1	1	2	6	3	3	4	3	1	3	2	....	....
45–49	....	2	2	....	....	1	2	4	1	....	2	....	1	2	2	2	....	6	2	1	1	2	1	1
40–44	....	....	1	....	....	1	1	2	....	....	1	1	1	1	....	2	....	1	1	1	....	1	2	4
35–39	....	1	....	....	....	....	2	1	....	....	....	....	1	....	3	....	1	3	2	....	....	1	1	....
30–34	....	....	....	....	....	....	....	3	....	....	....	....	....	....	....	2	....	....	....	....	....	1	....	1
25–29	....	1	1	....	....	....	....	2	....	....	....	....	....	....	....	....	1	2	....	....	....	1	....	....
20–24	....	....	....	....	....	....	....	2	....	....	....	....	1	....	1	....	....	....	....	....	....	....	....	....
15–19	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	1	....	....
	126	95	80	52	45	65	75	52	18	99	63	100	37	23	47	96	98	55	51	92	112	48	54	41
Average..	99.9	89.5	87.8	98.0	91.5	91.3	88.7	61.8	90.3	91.0	98.9	100.1	82.6	76.4	79.3	86.6	88.7	67.5	81.3	91.9	92.5	72.1	80.1	82.7
Average Mental Age	13^{10}	13^{0}	12^{10}	13	13^{2}	13^{1}	12^{11}	10^{8}	13^{0}	13^{1}	13^{9}	13^{10}	12^{5}	11^{10}	12^{1}	12^{9}	12^{11}	11^{2}	12^{3}	13^{2}	13^{3}	10^{8}	12^{2}	12^{5}

variation through a reclassification of the pupils should make the work of teaching very much easier and more effective. In another study we found that the variation in mental age was much greater than the variation in chronological age. This means that there is a tendency to allow children to progress in school in accordance with chronological age rather than mental ability, and this partially explains the wide variation of scores just noticed.

But the fact we wish to emphasize especially is the extent to which schools differ in the type of their pupils. This is better shown by a comparison of intelligence quotients. Scores will vary with growth, but the intelligence quotient is presumed to be nearly constant over a considerable number of years. Because of this constancy, the intelligence quotient is used to classify children as dull, normal, bright, or superior.

In Table II is shown the percent of pupils above and below certain intelligence quotients. Compare, for example, Schools 1 and 18. In School 1, 75 percent of the pupils have intelligence quotients of 100 or above; 57 percent of 110 or above; and 25 percent of 130 or above. In School 18 only 22 percent have intelligence quotients of 100 or above; 9 percent of 110 or above; and none of 130 or above. School 12 has children somewhat brighter than School 1; School 8 has duller children than School 18. Between these extremes are all gradations.

An examination of the data of Tables I and II shows that the school averages do not conform to the normal probability curve. This is because the tables do not include as many schools from congested districts as from the better residential sections. Were all schools included, there would be a larger proportion with low averages and the variation would be increased. Furthermore, in none of the schools are the extremely dull included in this grade, for they have failed of promotion and are either to be found in the lower grades or in special schools.

A comparison of these facts with the type of community served by each school is instructive. Following are brief statements descriptive of each community:

School 1—High-class residences; a large proportion of successful business and professional men.

School 2—In a congested district; includes stores and some factories; some foreigners; almost entirely a laboring class.

School 3—Mainly residential, but near factories; laboring class forms a large proportion of the population.

TABLE II. PERCENT OF PUPILS ABOVE AND BELOW CERTAIN INTELLIGENCE QUOTIENTS

Intelligence Quotients	Schools																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
100 or above	75	54	44	71	60	55	60	19	62	56	71	77	54	39	40	54	54	22	55	60	54	25	39	39
Below 100	25	46	56	29	40	45	40	81	38	44	29	23	46	61	60	46	46	78	45	40	46	75	61	61
110 or above	57	33	21	56	38	28	43	6	29	36	56	61	35	22	32	38	41	9	24	34	33	10	22	24
Below 90	11	32	34	19	18	22	23	62	25	20	13	15	30	39	43	31	27	64	33	25	27	50	39	34
130 or above	22	5	3	13	11	6	7	0	15	12	21	24	8	0	6	5	6	0	8	7	11	0	4	0
Below 70	0	3	5	0	2	3	9	23	2	1	6	1	5	9	13	5	3	15	8	3	2	8	7	12

School 4—High-class residences; homes of college professors and professional men; a small part is very wealthy and another part is decidedly tenement.
School 5—An outlying district; high-class residences; homes of business and professional men.
School 6—Good residences; homes of business and laboring men; an orphan asylum is within the district and sends its children here.
School 7—Good residences; a great number of excellent apartment houses; homes of business and laboring men.
School 8—Colored children only; a very high-class colored neighborhood; entirely residential; a great many apartment houses.
School 9—Good residences; many home owners; largely business men.
School 10—Thickly settled residential neighborhood; business and laboring men.
School 11—Good residences; apartment houses; business and office men.
School 12—Rapidly growing community; successful business and professional men; high-class residences; very homogeneous.
School 13—Outlying district; almost in an agricultural region; a very stable population.
School 14—Good residential; many negroes; laboring class.
School 15—Semi-residential; near the railroads and river; largely laboring class.
School 16—Part of district is high-class residential, part is tenement; an orphan asylum is in the district.
School 17—Rapidly growing district; mainly residential; homes of factory workers.
School 18—Almost in the heart of the city; very congested; low-class tenements; some foreigners; run-down neighborhood; laboring class.
School 19—Outlying district; has a great many children from an agricultural neighborhood.
School 20—Outlying region; high-class residential; business and professional men.
School 21—Outlying region; very good residences; business and working class.
School 22—Semi-residential; part of district is tenement; foreigners; laboring class.
School 23—Outlying district; fair residential somewhat rural.
School 24—Congested tenement district; run-down neighborhood; partly industrial.

It is manifest that these schools represent widely varying conditions. Children attending School 18 have a very limited home environment and nearly two-thirds of them are classified on the intelligence scale as dull. Children attending such schools as 1, 4, or 12 are of a radically different type both as to environment and general intelligence, being much above the average. These varying conditions indicate that the course of study is a problem for each school, to be solved according to local needs.

Enough experimental work has already been done to suggest how this may be carried out. Whipple and Terman have shown that bright children not only can make rapid progress but can at the same time take an enriched curriculum. It has also been found that dull children can almost never make normal progress through the grades. We thus have three groups, the dull, the normal and the bright, each requiring some special adaptation of the curriculum.

We have shown that these schools varied in the general intelligence of their pupils and that they served communities of very different types. It is also to be noted that there is a correlation between the intelligence level of the pupils and the character of the community. Those with lower intelligence levels generally live in the more congested districts; those with the higher intelligence levels generally live in the outlying, residential sections.

This suggests how the school may adapt itself to its community. Consider, first, how such a school as No. 18 will plan its curriculum. It is evident that it will have to permit its children to progress more slowly than usual, possibly allowing three years to do two grades of work. It should provide some of those things ordinarily found in the home but lacking in most of these homes, such as games, parties, plenty of simple reading, chance to "tinker," etc. Hand work should be provided, partly for cultural reasons, but mainly as a preliminary to vocational training. Since these children can neither make the rapid progress of bright children nor progress as far in school, one would only cause them to fail by urging them to meet the more difficult requirements. Failure tends to breed discouragement and discontent. If, on the other hand, the curriculum is within their capacity and definitely helps them to be self-supporting and to fill a place in society, their self-respect will be maintained and they can become valuable citizens.

Schools having bright children can expect the home to care for much of the training. These children can do more than average work, both in quantity and quality. They can possibly do three grades of work in two years and at the same time have their curriculum enriched. We do not know as yet just how the work should be enriched. A large proportion of such children will enter the professions or fill responsible positions in the business world. These vocations need a very broad foundation and an

extended amount of training. Anything the school can do to give better preparation in less time will be of great service to the individual as well as to society.

It is not likely that any school will be made up of but one of our three groups. It will generally have all three of them. An analysis of the kind here attempted will show which groups are present and what kind of adjustment is necessary to make the school of most service to its community.

SOME FURTHER STUDIES OF GIFTED CHILDREN

ELIZABETH CLEVELAND

Supervisor of Girls' Activities, Detroit Public Schools, Detroit, Michigan

Having established enough seventh- and eighth-grade classes for gifted children to serve most of the city of Detroit, and having arranged a satisfactory system of testing those who enter and of following up those who leave, we feel we have made a worth while beginning in the special training of the most promising. But we are still looking with curiosity and awe at the impressive group we have brought together, trying to discover "how they got that way" and what we had better do about it.

In the months between March and June, 1920 we made some special studies of our three "special advanced" classes, comparing them with a control group of normal pupils in the same schools as to health, nationality, home conditions, types of reading and recreation, amount of travel, and vocational and educational plans.

The health studies were made with 140 pupils and with an equal number of control pupils of the same age in the same schools. Although we were unable to arrange for a thorough physical examination of each pupil, we could at least, with the assistance of our physical training department, measure weight and height and make a rough estimate of general physical condition. Of course, compared with other pupils of the same grade, pupils of the special group are likely to be smaller, because they are younger. Many of them come from homes where health is intelligently looked after, and they show the effect of refreshing sleep, proper food, fresh air, and exercise. In comparing a group of gifted children with a normal group we had always received an impression of more complete physical fitness in the former. This is expressed in posture, in nervous control, and in a general look of contentment and well-being. There are fewer wandering eyes, fewer open mouths, fewer restless hands. Even taking into account the interest in work specially and skillfully adapted to their ability, these children show a poise and alertness that seem to be due, at least partly, to physical causes. According to the standards furnished by the Bureau of Education at Washington,

we found 114 of the special group against 82 of the control group to be within ten pounds of the proper weight for their height. The teachers' estimates of general health showed no striking differences between the groups. This evidence, so far as it goes, seems to refute the old idea that the brilliant mind is usually found in the unhealthy body.

In questioning the pupils as to nationality we went back to their grandparents. For 114 children in the control group there were 176 reports as to the nationality of grandparents and in the special advanced group 222.

The distribution of nationalities is shown in Table I.

TABLE I. NUMBER OF GRANDPARENTS OF INDICATED NATIONALITIES

Nationalities	Control Group	Gifted Group
American	97	90
English	13	28
Canadian	27	15
German	12	36
Scotch	17	23
French	3	11
Irish	5	16
Swiss	1	..
Polish	1	..
Swedish	..	1
Welsh	..	1
Newfoundland	..	1
Total	176	222

This report is, of course, not a safe basis of generalizing, yet it suggests some interesting lines of investigation. The English, Scotch, and Irish usually make a good showing and some teachers have attributed this to the common language. But the French and German make a still better showing and the Canadians not so good. In a general way we have always observed that the children of some nations are bright and some dull; that Russian-Jews are quick to learn and that Poles are slow. But what determines the fiber of a group that has lived for centuries in the same conditions? Is it the inner structure or the outward circum-

stances? If we knew we might begin to modify in some respects our complacent uniformity.

In another connection when studying a group of girls retarded for various reasons (other than mental deficiency) we had found the most marked difference from the control group in their fathers' occupations, indicating social rather than mental or physical causes for their retardation. Furthermore our first groups of gifted children, selected according to the judgment of teachers and principals, had shown a decided majority of pupils whose fathers were in responsible business positions or in the professions —pupils who belonged to the prosperous, or at least to the comfortable classes. Perhaps their good English, easy manners, and general sophistication impressed their teachers as superior intelligence. At any rate we were surprised to find that in the present study this distinction tended to disappear. When the gifted children were selected according to the results of the tests the occupations of their parents were chiefly remarkable for their great diversity.

When it came to tastes and standards of living, as shown in the kind of reading and recreation and in plans for education there was a distinct difference between the groups. The pupils were asked to name their favorite books, which were classified as inferior, average, and superior. Dime novels, silly sentimental tales, the "Elsie" and "Polyanna" sort of books were considered inferior. Books read for information, books that appeal through subject rather than style, ordinary modern novels, were classed "average." Well-written history, poetry, the finest fiction, were called "superior." The results are shown in Table II.

TABLE II. DISTRIBUTION OF PUPILS ACCORDING TO THE READING PREFERENCES

	Control Group	Special Group
Inferior	49	10
Average	55	54
Superior	9	49
Total	113	113

It seems obvious enough that they appreciate because they are bright. But can it also be said that they are bright because they appreciate? Would they have passed the tests as they did had they not read so much and so discriminatingly? Certainly the brighter ones even at this age have the vision of higher education. Sixty-two of the gifted group were already planning for college. Perhaps also the intelligence is sharpened by experience of the world. Forty-eight of the gifted group against 36 in the control group had travelled considerably.

At the end of the year a report was made on the pupils who were found in any way unsatisfactory, with the teachers' comments on the cause. We found 22 out of 160 reported weak, 15 of these in one subject only, 14 of these 15 in Latin and 13 of them in one school. Fifteen of the 22 were in the VIIB (the beginning class), six in the VIIA, one in the VIIIB, and none in the VIIIA. Only four pupils were reported generally weak. The main cause of the difficulty in the case of 11 pupils was reported as lack of concentration and application, which of course need not indicate that the testing was at fault. The problem is to develop these powers.

There were only four whom the teachers considered lacking in ability. These will be referred back to the testing department for special study.

A follow-up report on 47 high-school pupils formerly members of gifted groups shows the following results:

Superior work (almost all 1's and 2's)	30
Satisfactory (2's or 2's and 3's)	13
Unsatisfactory (3's or 4's in more than one subject)	4

A report from a Latin class, six of whose members were formerly in gifted groups and most of whom had had the same teacher since the beginning of their course, showed the following averages as results of their mid-semester examinations.

Average of class	66
Average of five from gifted group[a]	80
Average of class exclusive of five from gifted group	62

All these reports are suggestive so far as they go, but much more observation and experiment is needed. We plan to continue the recording along the lines indicated above and particularly to extend our work in physical measurement. Our first

[a]One best student absent.

classes are just graduating from high school and we hope to follow the progress of each individual in college or in vocations. So that while we have made no progress in the remoter problems of determining causes, we can at least begin to measure the effects of our methods and make modifications accordingly.

MOTIVATED DRILL WORK IN THIRD-GRADE ARITHMETIC AND SILENT READING[1]

J. H. HOOVER
State Teacher's College, Cape Girardeau, Missouri

THE PROBLEM

The problem of this study concerns the value of certain games or devices, to be explained later, in providing motivated drill work in the fundamental processes of arithmetic and reading. The study is based upon a few fundamental conceptions as to the nature and function of drill, i.e., of that kind of repeated activity which has for its purpose the increasing of one's physical skill or dexterity, or the permanent fixing in memory of certain useful associations. Drill, therefore, is an activity which has for its purpose the reducing of certain mental or physical operations to an automatic basis.

If drill is to be made effective and economical, it must be freed from some of its monotonous and unattractive aspects. When children see the need of the process, it will be interesting and impelling. This suggests utilizing the play instinct as evinced in games and dramatization. When drill is thus conducted it becomes something more than repetition; it becomes repetition *with attention*. Moreover, it addresses itself not merely to the group but to the individual. It is our purpose to show how this play instinct was utilized for drill purposes in arithmetic and reading and to indicate the results that were secured.

ARITHMETIC MATERIALS

In order to realize this purpose, suitable material had to be devised. In arithmetic the drill work assumed the form of a game in which the four fundamental processes were separately involved. After a number of plans had been tried and rejected, the one finally adopted involved the preparation of sets of cards two inches by one inch. On these cards numbers were written according to the arrangement in the case of dominoes. For addition, subtraction, and multiplication a set consisted of 28

[1] This investigation was carried on with the cooperation of F. J. Kelley, Dean of the School of Education, University of Kansas.

cards displaying the combinations of a set of dominoes which extends to double 9, except that all combinations containing 0 and 1 and all doubles were omitted. A diagram of this set follows. The letter "A" merely serves to identify a particular set.

A 2/3	A 2/4	A 2/5	A 2/6	A 2/7	A 2/8	A 2/9	A 3/4	A 3/5	A 3/6	A 3/7	A 3/8	A 3/9	A 4/5	
A 4/6	A 4/7	A 4/8	A 4/9	A 5/6	A 5/7	A 5/8	A 5/9	A 6/7	A 6/8	A 6/9	A 7/8	A 7/9	A 8/9	

FIG. 1. CARDS USED IN ADDITION, SUBTRACTION AND MULTIPLICATION. CALLED SET A FOR REFERENCE CONVENIENCE

Since the cards shown in Figure 1 are not adapted to short division others were devised for that purpose. The aim was, that the numbers and combinations of numbers, should be exactly divisible by as many of the numbers below ten as possible. Accordingly, two different kinds of card sets were devised called Sets B and C. The number 35 and its first 9 multiples formed the basis of the first set; and the number 72 and its first 9 multiples formed the basis of the second. For drill in division by 5 or 7,

A 35/70	A 35/105	A 35/140	A 35/175	A 35/210	A 35/245	A 35/280	A 35/315	A 70/105	A 70/140	A 70/175	A 70/210
A 70/245	A 70/280	A 70/315	A 105/140	A 105/175	A 105/210	A 105/245	A 105/280	A 105/315	A 140/175	A 140/210	A 140/245
A 140/280	A 140/315	A 175/210	A 175/245	A 175/280	A 175/315	A 210/245	A 210/280	A 210/315	A 245/280	A 245/315	A 280/315

FIG. 2. CARDS USED IN SHORT DIVISION WHEN DIVISOR IS 5 OR 7. CALLED SET B FOR REFERENCE CONVENIENCE

the set based upon the number 35 is used and for drill in division by 2, 3, 4, 8, or 9 the set based upon the number 72 is used. If we attempt to use but one set of cards basing it upon 2,520, the

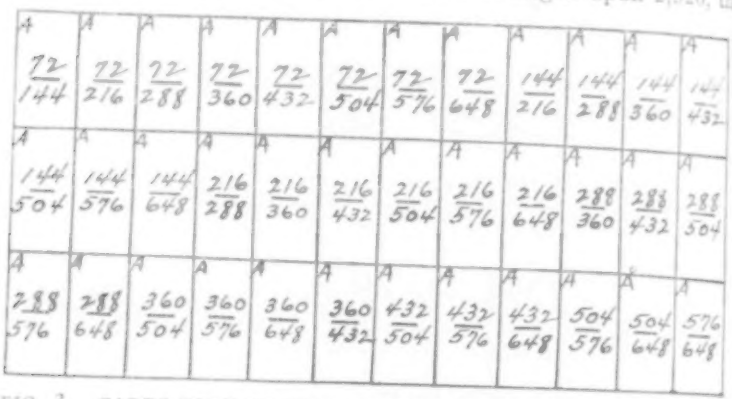


FIG. 3. CARDS USED IN SHORT DIVISION WHEN DIVISOR IS 2, 3, 6, 8, OR 9. CALLED SET C FOR REFERENCE CONVENIENCE

least common denominator of 2, 3, 4, 5, 6, 7, 8, and 9, the numbers are entirely too large for third-grade children to handle. Figures 2 and 3 show the cards as adopted. It would be better if the numbers were smaller and but one set were used; but to bring about this degree of simplification seems to be impossible.

Reading Materials

As in arithmetic, the aim in devising reading materials was to have drill work in reading assume the form of a game. In this, the elements of comprehension and speed were to play a prominent part.

Since love of activity is one of the characteristics of childhood, this fact was used in developing the reading materials. Printed cards containing "action" sentences which lend themselves readily to dramatization in the school room were devised. In determining the content of the sentences, the environment, interests, and every-day activities of children as a whole were kept constantly in mind. In order to appeal to the needs and interests of various types of children, four different kinds or sets of cards were devised which, for convenience, we will call, Sets A, B, C, and D. For a

description of these different sets of cards see, "Rules for the Reading Game," page 206.

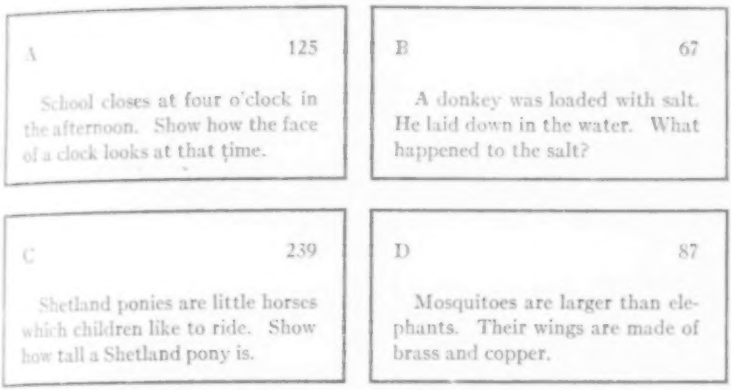

A 125
School closes at four o'clock in the afternoon. Show how the face of a clock looks at that time.
B 67
A donkey was loaded with salt. He laid down in the water. What happened to the salt?
C 239
Shetland ponies are little horses which children like to ride. Show how tall a Shetland pony is.
D 87
Mosquitoes are larger than elephants. Their wings are made of brass and copper.

FIGURE 4. SAMPLE CARDS IN READING

A sample card from each of the four sets of reading cards is shown in Figure 4. Each card is 2 inches wide and 4 inches long. The A, B, C, or D, as the case may be, which appears on each card indicates the set to which the card belongs. There are 150 A's, 150 B's, 250 C's, and 100 D's. The cards in each set are arranged in order of difficulty (least difficult first, most difficult last), and the number of the card indicates its position in the set. For instance, A:125 means that the card belongs to Set A and is the one hundred and twenty-fifth card (based upon the author's judgment) in Set A from the standpoint of difficulty.

Set A is a group of "Action Cards." These cards are primarily simple commands or requests. The child works with things actually present, and no pretense is involved. Bodily activity is required in each case—e.g., place your right hand on your left knee.

Set B is a group of "Language Response Cards." Response to these cards is made wholly through the medium of spoken or written words. (Name some good winter games.) Bodily activity is not required. Language responses may be written or oral, depending upon the teacher's judgment as to the needs of the particular group of children in question.

Set C is a group of "Pretense Cards." Here the children are asked to *pretend* that they are doing this or that particular thing.

They work, or pretend to work, with things *not* actually present. (Act as if you were hoeing in the garden.) Muscular activity is required in all cases.

Set D is a group of "One Word Response Cards." Response to these cards may be made by using one of the four following words: *yes*, *no*, *right*, or *wrong*. For example, to the question *Is ten greater than nine?*, the response is *yes*. Again to the statement *Horses have two feet*, the response is *wrong*.

If a group of children need exercise in giving correct oral or or written language responses they should be given cards from Set B. If they need exercise in accurately getting the thought from the passage read so that they can perform the desired activity and thus give visible evidence of understanding or misunderstanding the passage read, they should be given cards from Set A or C. If they need practice in selecting the correct answer where other answers are possible, they should be given cards from Set D.

Rules for the Arithmetic Game

Addition.—Set A (numbers below 10) is used. The rules for playing this game are very similar to the rules for playing dominoes. Being adapted, however, to the needs of third-grade pupils. Children will learn to play the game more quickly if the teacher plays a game with one of the pupils and allows the others to watch while it is being played.

The pupils are arranged in pairs according to some convenient plan; for example, the child in the front seat plays with the one behind him, etc. Each pair of pupils is given a package of cards. For convenience let us say that Ruth and James are playing together. Each has pencil and paper on which to keep his own score.

James lays the cards on the desk, face downward, and shuffles them. They draw out seven cards apiece. Ruth holds her cards so that James cannot see them and vice versa. Ruth lays a card on the desk (say $\frac{4}{5}$) and adds the two *end* figures together. She puts "9" on her paper as her score for this play. From the seven cards in his hand James matches this card. Any card containing a 5 or a 4 will do (say $\frac{4}{7}$). Adding the two *end* figures (5 and 7) together he gets 12 as the sum, and puts "12" on his paper as his score for this play. Thus the game proceeds. If at any time Ruth cannot match the cards on the desk from the cards

in her hand she loses her chance to play and James plays two (or more) times in succession. Ruth plays again as soon as she can match the cards on the desk from the cards in her hand. There is no drawing from the unused cards. The game ends when one of the players no longer has any cards in his hand. They then add their scores, and the one having the *higher* total score wins the game.

The used cards are turned face downward and shuffled with the unused cards, then a new game begins. Playing is continued in this way until the time is called. Promptness on the part of the pupils in beginning and ending the play period will add much to the interest and usefulness of the game.

Subtraction.—Set A is used. The rules for "addition" apply to "subtraction" also, with the following exceptions; this time the end figures are *subtracted* (the smaller from the larger) instead of added. For the plays given in the above illustration, Ruth places "1" on her paper for her first score instead of "9" and James gets "2" instead of "12." At the end of the game the results are again added but this time the one getting the *lower* total score wins the game.

Multiplication.—Set A is used. The rules for "addition" also apply to "multiplication" except that this time the end figures are *multiplied* instead of added. Again referring to the illustration, Ruth places "20" on her paper for her first score and James gets "35." At the close of the game the scores are again added and the one getting the *higher* total score wins the game.

Division.—Set B (multiples of 35) or Set C (multiples of 72) is used. Sets B and C are very similar in purpose. They are designed for drilling pupils in the addition, subtraction, and multiplication of numbers *above* ten but they are *especially* designed for drilling pupils in short division and, to a limited extent, in long division. When the divisor is to be 5, 7, or 35, Set B is used. When the divisor is to be 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, or 72, Set C is used.

The rules for playing differ in no fundamental way from those already given for addition. Children are arranged in pairs. Cards are shuffled, etc. For this particular day, say, drill in dividing by "4" is desired. The teacher writes the figure "4" on the blackboard where it can be seen by all. Again using our illustration; Ruth lays a card on the desk (say 72/288). Here the

end figures can either be *added* or *subtracted* depending upon the teachers' judgment as to which the class most needs—drill in addition or subtraction. In either case the result is *always* exactly divisible by four. If the *end figures* are *added* (using illustration above) the problem will be, 4 into 360, and Ruth will put "90" on her paper as her first score. If the *end figures* are *subtracted* the problem will be, 4 into 216, and the result is 54. The game proceeds until one of the players no longer has any cards. The scores are added, the one with the *higher* total score winning the game.

If drill is desired where the divisions are *not* always exact, use Set C when divisors are 5, 7, or 35; and Set B when divisors are 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, or 72.

Rules for the Reading Game

The children are arranged in pairs according to some convenient plan. Each child is given a sufficient number of cards to occupy his time for the entire reading period. If the time allotted to a reading period is fifteen minutes, ten cards given to each child will probably be enough.

Suppose that Ruth and James are playing together. Each is given (say) ten cards from Set A. Each has a pencil and paper on which to keep the score of his or her opponent. James picks up one of his cards, reads it silently, and hands it to Ruth who also reads it. He then proceeds to perform the required activity. By his performance, Ruth judges whether or not he has the thought of the passage he has just read. She now gives him a score of "1," if he has performed his task correctly or of "0" if he has failed.

The teacher will do well to be in the midst of the children while the game is in process. She should watch the performances of the children who are being judged and the scoring of those who are doing the judging. Fairness, accuracy, and speed are to be encouraged.

Ruth now reads one of her cards and James becomes judge. Thus the game proceeds until the twenty cards are exhausted or until the reading period has ended. The one having the greatest number of perfect scores at the end of the play period wins the game.

The rules for Sets B, C, and D are the same as those for Set A. The only difference is in the nature of the response, and this does not affect the rules for playing.

Experimental Procedure

The experiment involving the use of these materials was carried out at Kansas City, Kansas. It included 1,139 third-grade children (571 in non-drill and 568 in drill sections) or thirty different third-grade rooms. The general plan of procedure was as follows:

1. At the beginning of the study (December 29, 1919) and at the end (April 1, 1920) standardized tests in arithmetic (Cleveland Survey Arithmetic Tests) and reading (Monroe's Standardized Silent Reading Tests) were given to all the children.

2. After the first tests had been given, an effort was made to divide the pupils of the thirty different rooms into two groups of equal size and mental attainments. In making this division the advice and assistance of the superintendent was sought. This method of division seemed advisable instead of waiting for the test results because it was desirable to begin the study without delay. The superintendent's judgment in this matter was exceedingly accurate; and this accurate division adds greatly to the value of the results of the study.

The division into groups accomplished; fifteen of the rooms were provided with the materials and the teachers were instructed in their use. The teachers of drill classes were requested to use the materials, both arithmetic and reading, 10 minutes a day on Mondays, Wednesdays, and Fridays. These instructions were carefully observed.

Since the aim was to compare the improvement of the drill group with that of the non-drill group, it was essential that the same amount of time be spent by each group in arithmetical or reading improvement. Therefore, the time spent in the extra drill work by the drill section was deducted from the regular amount of time given to arithmetical or reading improvement. In other words the time element in the two groups was identical, the only difference being the way in which this time was utilized.

In arithmetic the time allotted to the experiment was, as nearly as possible, equally divided between the four fundamental processes. No special instructions were given to the teachers of drill classes in arithmetic and reading (aside from the printed rules) except that they were to emphasize both speed and accuracy in the four operations of arithmetic and speed and compre-

hension in reading. They were to cover the same daily assignments in the textbook as the classes that did not use the drill materials. The teachers of the non-drill classes were asked to proceed with their classes in their usual manner, making no changes in their methods of instruction.

Methods of Scoring Papers

All the test papers were scored by the writer according to uniform methods. These methods followed the suggestions of the authors of the tests, except that in the case of the Cleveland Arithmetic Test, for which no accuracy score was provided, the writer used his own method of recording accuracy. This consisted simply in finding for each pupil the percent of correct answers on each sub-test, and in computing the mean of these percents.

The accuracy score for an entire class was obtained by adding together the accuracy score obtained by each class member, then dividing this sum by the number of pupils in the class.

For the methods of scoring speed in arithmetic, and rate and comprehension in reading, the reader is referred to the directions for scoring these items as found upon the score sheets which accompany these tests.

Results of the Experiment

The improvement of each section (drill and non-drill) as shown by the tests was calculated and put in tabular and graphical form (not shown in this article).

In every phase of arithmetic and reading considered in the study, the improvements of the drill section were more pronounced than the corresponding improvements in the non-drill section.

Let us first consider arithmetic. In Test A of the Cleveland Survey Test, the non-drill section changed during the study from a median of 10.9 to a median of 15.1 examples correctly solved in 30 seconds.

The corresponding improvement in the drill section was from a median of 9.7 to a median of 18.5 examples correctly solved in 30 seconds. Here the gain is 8.8 examples as compared with a gain of 4.2 examples in the non-drill section.

Similar comparisons of the gains (Tests A, B, C, D, E, F, G) made by the two sections (drill and non-drill) may be obtained from Table I.

TABLE I. COMPARISON OF ARITHMETIC GAINS MADE BY THE TWO SECTIONS

	Tests							Accuracy
	A	B	C	D	E	F	G	
Drill section	8.8	5.7	6.1	7.0	1.9	1.7	1.2	17.8%
Non-drill section	4.2	3.6	4.4	4.5	1.1	1.3	1.0	14.1%

In studying the subject of class gains two factors should be considered: (1) advancement along the scale of measurement of the median performance, and (2) the increase or decrease of class variability.

By referring to Table II the reader will note that in a large majority of the tests the *relative* variability in both sections (drill and non-drill) decreased during the study but that the decrease was more pronounced in the drill section than in the

TABLE II. COMPARISON OF ARITHMETIC VARIABILITIES IN THE TWO SECTIONS

A. NON-DRILL SECTION

Tests	January			April		
	Median	Quartile Deviation	Coefficient Variability	Median	Quartile Deviation	Coefficient Variability
A	10.9	2.6	0.24	15.1	3.1	0.27
B	5.2	2.3	0.44	8.8	3.1	0.36
C	2.9	1.0	0.35	7.3	2.3	0.33
D	3.0	1.1	0.37	7.5	2.5	0.33
E	2.6	0.8	0.31	3.7	1.1	0.30
F	0.7	0.4	0.59	2.0	1.2	0.60
G	1.1	0.5	0.45	2.1	0.9	0.43

B. DRILL SECTION

Tests	January			April		
	Median	Quartile Deviation	Coefficient Variability	Median	Quartile Deviation	Coefficient Variability
A	9.7	2.6	0.27	18.5	3.7	0.20
B	5.0	2.2	0.44	10.7	3.6	0.34
C	2.7	1.0	0.37	8.8	2.6	0.30
D	2.9	1.0	0.35	8.9	2.7	0.30
E	2.5	0.8	0.32	4.4	1.3	0.30
F	0.76	0.5	0.66	2.5	1.1	0.44
G	1.2	0.5	0.42	2.4	0.8	0.33

non-drill section. It will be noted also that the respective increases of *absolute* variability in the two sections are about equal. For example, in Test G the non-drill section changed from a quartile deviation of 0.5 and a variability coefficient of 0.45 at the beginning of the study to a quartile deviation of 0.9 and a variability coefficient of 0.43 at the close of the study. In the drill section the corresponding changes were from a quartile deviation of 0.5 and a variability coefficient of 0.42 to a quartile deviation of 0.8 and a variability coefficient of 0.33.

In reading greater improvements were also realized in the drill section. In comprehension the drill section gained 5 units as compared with an increase of 3.1 units in the non-drill section. In rate of reading the drill section gained 21.2 words per minute as compared with a gain of 12 words per minute in the non-drill section. Table III shows the details.

TABLE III. COMPARISON OF READING GAINS MADE BY THE TWO SECTIONS (MEDIANS)

	Comprehension		Rate	
	Non-Drill Section	Drill Section	Non-Drill Section	Drill Section
January.....	3.9	4.0	34.7	35.7
April.....	7.0	9.0	46.7	56.9
Gain.....	3.1	5.0	12	21.2

In rate of reading the absolute variability (average deviation) increased for both sections, but the increase was larger for the drill section. The relative variability, however, owing to the larger gains in median performance was smaller at the end of the period than at the beginning. In comprehension of reading there was a slight decrease in absolute variability for the non-drill section. Both sections showed a decrease in relative variability. These facts are shown in Table IV.

TABLE IV. COMPARISONS OF VARIABILITIES IN READING

	Average Deviation			Coefficient of Variability		
	January	April	Increase	January	April	Decrease
RATE						
Non-drill section.	3.3	3.7	0.4	0.84	0.53	0.31
Drill section.....	3.1	4.5	1.4	0.78	0.50	0.28
COMPREHENSION						
Non-drill section.	19.4	19.0	—0.4	0.56	0.41	0.15
Drill section.....	17.7	21.0	3.3	0.5	0.37	0.13

Suggestions Afforded by the Study

1. Uneconomical methods of drill are now being employed in the lower grades of our public schools.

2. Greater use should be made of the doctrine of interest, especially as it applies to drill work. Drill work should be motivated or vitalized by being connected with some dynamic purpose.

3. Bodily activity (dramatizations, handling of objects, etc.) can be profitably connected with achievement in school subjects.

4. Drill, to be efficient, must be made individual in character. It should be conducted, as nearly as possible, according to a child's needs and particular abilities.

5. Intensive focalization in connection with attentive repetition is an essential characteristic of efficient drill work, and by appealing to the play instincts of children this desired characteristic is effectively provided.

SOME ELEMENTARY STATISTICAL CONSIDERATIONS IN EDUCATIONAL MEASUREMENTS

J. Crosby Chapman
Yale University

The determination of norms of achievement for educational tests is admittedly a matter of great labor. It is the privilege of intelligence to analyze the exact advantages which are derived from any activity, especially when the activity degenerates into monotonous toil.

The question as to what are the exact uses to which norms of achievement are put can only be answered by considering the more general question. For what precise purposes are standard tests administered to a group? Two purposes are at once apparent: (1) to determine roughly the extent of the individual differences within the group; (2) to compare the achievements of the individuals of the group with those of other individuals external to the first group.

Obviously for the first use norms of achievement are superfluous. The differences within the group are the point of interest. Ninety-nine times out of every hundred, when a test is applied the major interest, if not the only interest, resides in the individual differences between the different members of the group. Even where an interest may extend to a comparison of the performance of this group with country-wide standards, such comparison is usually of little avail for reasons to be stated later in the article. In many cases we may be sure, therefore, that it is time wasted to secure norms based on large numbers. No tests which have been given sufficient trial to insure their validity should be held back because of lack of norms. The first point which demands consideration before attempting to collect standards on any test is: Is this test in its present form worth the time and energy that will be needed to collect a reasonably accurate norm? No physicist would dream of determining the constants of an instrument with great care, unless he was reasonably sure that the instrument was to be put to uses where such precise knowledge of constants would be required.

First, let us consider the question of the determination of norms and later we may take note of some considerations with reference

to construction of the scales themselves. It seems to have been accepted without thought by a number of educational workers that the values derived from the application of a test to several thousand children are of necessity more accurate and more valuable than the results obtained from a smaller group. There exists a pathetic trust in the saving power of large numbers. It is apparently thought that a norm is made much more accurate by increasing the number of cases from one thousand to seven thousand, and some authors have gone so far as to quote with pride that the norms are based on more than sixty-eight thousand cases!

It is fairly easy to see how the misconception has arisen that a mere increase in the number of cases must, of necessity, yield more accurate results. Our knowledge of the law of error has been evolved from a theory of probability, based on ideal conditions. We all have a nodding acquaintance with the perfectly homogeneous disc or coin, which mathematicians toss, and with the urns from which are drawn the gayly colored balls. Under such conditions as those found in the tossing of the coin, for example, it is obvious that as the number of cases is increased, the ratio between the number of heads and the number of tails tossed becomes increasingly nearer to unity, which we judge by experience is the ideal value to which the ratio is gradually approaching.

The reason why any considerable increase in the number of cases gives a more accurate value of this ratio is found in the fact that the successive tosses are performed under a definite and limited set of conditions. The first five tosses, we may say, form part of an infinite series, the series being homogeneous in the sense that each member expresses the result of a toss, performed under conditions which do not change. With such series, any considerable increase in the number of the series tends to give a more accurate value for the ratio which is being determined. There is no limit to the approach of this ratio to unity. The series can be increased until human patience or human effort is exhausted. Every addition to the series tends to increase the accuracy, though after a certain point is reached the increase in accuracy becomes almost negligible. Increasing the length of the series is theoretically valid when the series is homogeneous, and is practically beneficial when the resulting increase in accuracy adds materially to the

value of the norm as a practical instrument, with a full consideration of the conditions under which it is to be employed.

To pursue the second point with reference to increased accuracy as a result of increase of cases, if the series fulfills the requirements of homogeneity, then the simple formula holds, and the probable error of the arithmetic mean of a series of observations is inversely as the square root of their number. This fact shows how easy it is to overrate the effect of multiplying observations. How neglible must be the added degree of accuracy, given by the addition of fifty-nine thousand cases to an original nine thousand cases! If the material is homogeneous, the above formula holds and the gain in accuracy is small compared with the experimental error incident to the collection of such data. If, on the other hand, the material is heterogeneous, the above formula is no longer applicable, and the general procedure of compounding becomes meaningless. The accuracy of the educational norm is not primarily controlled by the total number of cases and the use of the probable error formula is to most readers very misleading; the value of the norm depends almost wholly on the care with which the selection of subjects has been made. It is with reference to this very question that the greatest carelessness has been shown by educational workers.

If we refuse to be blinded by the large number of cases on which the standards of achievement are based and examine the conditions under which this mass of material has been "rounded up," the confidence in the reliability of the norm declines. Taking any one of the older tests, there is no doubt that the times at which the test was applied to any particular grade have varied all the way from October to June. Grade VI is of course Grade VI. whether it is measured in October or June, but the results obtained in October cannot be compared with those of June. We might suppose that when we run into thousands of cases, the final norm would at least give us the average result of February. But two assumptions are made here. The first is that the curve of improvement is strictly linear between October and June, which we know is not the case; and the second is that the tests have been equally distributed over the entire school year. But tests are usually employed to measure the attainments as a result of certain instruction. This means that more of the tests are given in the later months of the school year than in the earlier months.

Or the problem can be looked at from another point of view. Suppose the absolutely true values of achievement for the whole of the United States (whatever that may mean) in a particular test on the last day of the school year are

Grade V	Grade VI
9	18

Then we may assume that the result of each month of instruction is to increase the efficiency by one unit. Let us suppose therefore that instead of the norms being determined on the last day, they have been spread over a period of the last month and a half, a generous assumption. Then the average attainment of Grade VI will be approximately 17¼ which differs ¾ from the true mean. Yet compilers of statistics are determining the value of these indefinite measures (in our case 17¼) by using thousands of cases! To go on piling up the number of cases in the hope of getting accuracy, when such a fallacy as the above exists, is futile. One might as well send to the Bureau of Standards for accurate electrical instruments to determine the resistance of a bar, and then allow the temperature to vary in an uncontrolled way anywhere from 20 to 100 degrees centigrade. Nothing further need be said with reference to the fallacy of selection and the actual discrepancies in administration; the first is a large factor, while the errors introduced by the second factor are sufficient to make compilers more guarded.

It may be urged, however, that all the members of a series upon which a grade-norm determination is based have at least this one factor in common, namely, belonging to Grade X. While this is of course true, the fact of belonging to Grade X is of little significance, seeing that it may mean, as we have shown, anything from member of Grade X in October to member of Grade X in the following June. Furthermore, Grade X has no magical constancy, even if we confine it to Grade X in June, or let us say the last week of the period spent in the grade. It is a notorious fact that grades vary enormously, even within the same school system. The variation becomes even greater when we include numerous school systems; and when we include the grades of the smaller school systems and rural schools, the factor of belonging to Grade X ceases to have any valuable significance. It seems a lamentable waste of time, when one essential element in the equation is this appallingly variable factor, to bother with one hundred thousand

cases, or even with the second thousand cases, in the hope of correcting the second or third decimal place in some other variable. As time advances, as educational procedure is altered, as the attitude to retardation and elimination fluctuates, the meaning of Grade X must always be unknown. The norms of educational measurement, whether by grade or by age, need only be determined in the very roughest manner, for determination by age is by no means free from blemishes!

It cannot be too clearly recognized that the combining of a large number of groups, each homogeneous within itself, into one large heterogeneous group, in order to determine a very significant norm, defeats its own purpose. The norm for each of the homogeneous groups may have been of interest and of value, but the combined norm is almost valueless. Thus, for example it is fairly significant to know that a particular test, tried out in a particular city on Grade VI children, during their last month in the grade, gave a median result of 18. But when these data are combined with the results of a very inferior school system and even of rural schools, the composite value becomes meaningless. Mongrel results of this kind are like mongrel animals, of little worth. Each of these values has significance, but when compounded they becomes noninterpretable.

The best plan therefore is to keep the distributions separate. It would be far better for any standard test to be accompanied by the actual distributions, determined at dates specifically stated, on groups specifically described, than to attempt to determine a median for one million cases made up of diverse groups. Even if we can assume that in future an attempt will be made to have the test standardized for the close of the school year, it will probably be beneficial to publish the norms, obtained from twenty or thirty representative schools, rather than to combine these norms into a single figure.

Another consideration which enters with the passage of educational measurements from its infancy is the factor that the results of a particular year cannot be compounded with those of the preceding year, unless nothing has happened during those years to alter the conditions of learning, etc. If new methods are being used, if the effect of giving tests has been to alter attitudes, to improve instruction, to increase the emphasis here, to eliminate the useless there, then obviously to combine the results is to

break the law which admits of such compounding. As Venn says, "There is a familiar, practical form of the same error. . . . It is that of continuing to accumulate our statistical data to an almost indefinite extent of time or space. If the type were absolutely fixed, we could not possibly have too many statistics. The longer we chose to take the trouble of collecting them, the more accurate our results would be. But if the type is changing, if, in other words, some of the principal causes which aid in their production have in regard to their present degree of intensity strict limits of time or space, we shall do harm rather than good if we overstep these limits." Thus, for example, to compound the norms which Ayres publishes for his spelling test with present-day norms would be to commit the above fallacy.

It may be urged that there can be no objection if a person wishes to determine what is the average median performance of the whole country. Certainly not, especially when the individual in question pleads academic interest in the national norms for their own sake. But we may reasonably ask, apart from the satisfaction of this peculiar interest, to what conceivable use this norm, acquired with great labor, can be put?

Very accurate work in the collection of norms is futile until the following details are known with regard to the members of the group: (1) chronological age; (2) mental age; (3) sociological status; (4) hours devoted to subject. While some of these facts can be quantitatively measured, others equally essential can only be roughly estimated. A norm can never rise above its origin. Shacks are suitable if one builds upon sand, but lofty edifices demand deep foundations. We must always remember that we are measuring human traits, the conditioning factors of which are most complex; and we must estimate our errors and modify our procedure accordingly. Until we have much greater insight into what constitutes random sampling, let us beware of wasting our time.

We may now consider another phase of educational measurement, the neglect of which is threatening the effectiveness and speed of growth of the test movement. Obviously the school situation demands tests and yet more tests, tests that are sufficiently accurate, not to reveal the acumen of the men who construct them, but for the practical uses to which they will be put. This in itself should stimulate the output of tests which provide rough estimates of traits that need measurement. One is slow

to blame a scientific worker because he is too careful, but he may be correctly blamed if he breaks his own scientific code. It is perfectly legitimate for the scientist to say, "I refuse to meet this practical demand," but he must obey certain well-defined rules or he ceases to be a scientist. One of these elementary principles of scientific method often overlooked is the important one of balancing the errors within an experiment. In one part of the experiment we must not attempt to get accuracy to one-tenth of a percent, when in another part of our experiment there is an obvious error of 3 or 4 percent. Educational statisticians are most guilty of this time-consuming, fallacious procedure.

An illustration of a failure to balance errors which has gone unchallenged may be taken from Trabue's valuable study of completion tests. It will be remembered that the author wades through an enormous amount of statistical calculation in order to determine with slightly greater accuracy the relative difficulties of the various sentences on his scale. When has his scale been used for a purpose which demanded this refinement? However, ignoring this practical consideration and confining ourselves wholly to the scientific technic involved, it may be pointed out that the assumptions which have been made by the author are so great that any attempt at extreme accuracy is absurd. Note a few of the assumptions made: (1) that language completion ability is represented by the much abused probability curve; (2) that the variability of the different grades is equal; (3) that the time allotment was sufficient to give a "reasonable amount of time for each sentence." One almost hesitates to think what would be the effect on the elaborate superstructure of individual values of each sentence had they been given with different time limits and in a slightly different order. In spite of these assumptions, the writer estimates his intervals between grades by three methods, working in each case to the third decimal place and in some cases actually to the fourth decimal place! Incidentally I may remark that my own work shows the same weakness. We have all sinned. Most of us have strained at gnats and swallowed camels. Possibly swallowing camels is the fate of the pioneer, but it is a pity for the output of tests to be small because of the gnats.

Similarly the calculation of regression equations for purposes of weighting, as in some of Kelley's recent work is often much more a measure of the degree of random sampling attained than it is a measure of the relative effectiveness of the various measures involved. As a method, the study is of value, but the results are reeds, painted to look like iron. Several recent articles show that there is danger of the partial correlation method degenerating into a fashionable fad rather than serving a limited practical purpose.

The same is true in our straight laboratory work. I well remember my surprise on being told by an instructor in a psychological laboratory when engaged in the star tracing mirror experiment, that I must use a stop watch and not my own watch in order to read to fifths or at least half seconds. I tried to point out that there was an error of a very high order in the reckoning of what constituted the necessity for a retrace, but it was all to no avail. Accuracy was necessary for its own sake.

In the majority of cases where correlation is employed, the fact of interest is usually whether the correlation is high or low, yet how often the author, with apparent desire for intellectual exercise for its own sake, will determine the relation by both rank and product-moment methods, when one would be ample for the purpose in view. In many cases if judgment had been used it would have been seen that the correlation formula itself could hardly be applied under the circumstances. A friend of mine, when engaged on a piece of work which involved about two hundred correlations, had occasion to show it to some advisors. He had used the quick foot-rule method of correlation (his data made the use of any formula rather doubtful), and in any case he was interested merely in big differences. Yet with one exception his advisors suggested to him that he use the more refined formula of product moments. On pressing them for ultimate purpose they all admitted that their reason for suggesting it was based on habit and not on analysis. I am afraid the publication of correlation data to the second place when the total interest is in the first figure, and where the assumptions and shortcomings of the experiment have nullified the accuracy of the first figure, let alone the second, has had a bad effect on the scientific conscience. No corrections for attenuation can balance limited or slipshod selection of cases.

I have not considered the time consumed by many writers in perfecting a test in one direction while in other directions there are glaring practical deficiencies. Nor have I stressed the fact that within my whole experience I have never met a situation where anything but a rough instrument was needed. Perhaps one of the causes of this false striving for refinement may be found in the fact that Columbia University, which has so ably led the way in this type of measurement, has produced its studies for the double purpose of training immature scientists in methodology and of meeting a practical demand. The first factor, combined with the eagerness of educators to establish the dignity of their new-born science, has resulted in the latter being dressed in the garments of the more refined sciences. Too many clothes impede an infant science. Great care must be taken that the training in accuracy is not accuracy for accuracy's sake, but balanced accuracy which contributes to the attainment of a scientific goal. Those who find pleasure in very refined measurements for their own sake are prostituting their talents in the pioneer work of educational measurements. For keen analysis there is crying need in education but for ultra-refined measurement there is a more imperative demand in other fields; in these more settled regions this rare inclination will find ample scope and great reward.

Editorials

VARIABILITY

A bit of description in a recently published short story contains the following passage: "He was a tall man, with a red face, a large nose, fat cheeks, and a pendulous lip. His arms were long, etc., etc." Perhaps this is enough of a rather disagreeable picture; and, as in the story, so here, we shall skip to the next paragraph.

Such a passage really does evoke a picture, if we stop to let it. It has meaning. We are led to inquire how this comes about. In particular, how red must a face be to be called red, or how fat may one's cheeks be and yet escape the epithet? Under what circumstances are lips pendulous or arms long? How large is a large nose and will the same dimensions when applied to an eye or an ear justify the same adjective? Or, to proceed to extremes, would they justify it if applied to a man? This, you will say, is absurd. A nose three inches long is undeniably large, but a man three inches long would be too small to be a man. Even a human hand three inches long would not be large, but, on the contrary, pitifully small. These things, though equal, are yet unequal. Though they have the same measurements, they are somehow large or small in virtue of something independent of their size.

The fact seems to be that each judgment we make is with reference to particular standards. Although we measure teachers' salaries and the cost of office furniture in the same units we judge salaries and furniture costs each by their own standards. Cheeks are fat and noses are large because of something we have organized into our experience about the fatness and largeness of these features. What is the character of this experience? What data does it provide? By what fusion of past impressions are we enabled to pass judgment upon objects and processes, upon abilities and characteristics, upon men and events, shaping in what we call a reasonable way our thoughts and actions with reference to them?

We maintain that the data by which these concepts develop are twofold. In the first place, we recognize consciously or otherwise a type or standard for each class of things; and in the

second place, we have a notion of the extent to which each thing may depart from its type or standard without becoming unusual and the extent to which it may do so without changing its class.

Now, there is a very real sense in which statistical method abbreviates this slow process of forming concepts. For the type or standard it offers the average or measure of central tendency; for the extent of departure from type it offers the measure of variability. In using statistical method, therefore, we simulate the natural processes of the mind. We gather, for example, measures of the heights of Anglo-Saxon men—men whom we have never seen and whom other people measure for us, vastly extending by this means the range and number of our observations beyond those which we could possibly make personally. By rigorous methods we determine the average height of these men, and we probably find it to be about sixty-eight inches. Owing to the large number of cases at our disposal and the precision of the method employed, this result undoubtedly gives us a far more accurate notion of the "man of average height" than mere fortuitous experience could ever afford.

But our judgment as to whether a man who is above or below average is tall or short, and as to whether he is moderately or excessively so, depends not only on how much taller or shorter he is than the average but also on the closeness with which the heights of all men are found to group around the average. What serves to distinguish an individual is not merely his deviation from the average, but in reality that deviation in relation to the *typical deviation* of men in general. In other words to know that a man is seventy-one inches tall and hence three inches above average does not permit us to call him a tall man unless we know that three inches above average is unusual. Now, as to what we have just called the "typical deviation," we should find on examining our data that half the men whose measurements we had secured were within about two inches of average height. Our line of thought, therefore, might be somewhat as follows: this man is three inches taller than the average; of men in general who are above average height, half of them exceed the average by two inches or less; therefore, this man is tall.

"Two inches" in this case is a measure of variability. In so-called normal distributions it is the median deviation—that deviation from the average which is exceeded in half the cases.

Applied to men of shorter-than-average stature (average being sixty-eight inches) it would suggest that those of less than sixty-six inches might be called short.

Of course, we may be so constituted that we withhold the adjective tall or short—or hot or cold, rich or poor, skilful or unskilful—unless the object possesses appreciably more or less than this average-plus-or-minus-the-median-deviation of the quality in question. This is a personal matter; and because it is so, we encounter that vagueness of meaning that so often interferes with the understanding of language. Nevertheless, it seems not unreasonable that, for example, a man over seventy inches (five feet ten inches) in height should be called tall while one less than sixty-six inches (five feet six inches) should be called short.

In the case of intelligence quotients, the average is 100. But much larger numerical deviations from the average occur than in the case of heights of men. The median deviation, instead of being something like *two*, is about *twelve*. This means that excess above (or defect below) the average must be numerically six times as great in order to be equally significant. In other words, *twelve* above average has the same meaning in respect to brightness that *two* above average has in respect to height; that *twenty-four* above average in the one case means the same as *four* above average in the other; that *thirty-six* corresponds to *six;* etc.

Since in this sense *twelve* units of brightness have the same meaning as *two* units of height, these amounts are given a common name (median deviation); and this median deviation is used as a common unit.

We may put this in schematic form as follows:

Average height of men = 68 inches
Median deviation (M.D.) = 2 inches
Average brightness = 100 I. Q.
Median deviation (M.D.) = 12 I. Q.

Assuming the accuracy of these figures, we may make certain statements, such as that a tallness of 72 inches (2 M. D. above average) equals a brightness of 124 points (also 2 M. D. above average), or that a tallness of 74 inches is three times a brightness of 112.

At first sight this sort of comparison may seem forced. Yet as a matter of fact it is very common. We lately heard it said of a man who smote an anvil on week-days, but who smote a pulpit

on Sundays to better effect, that he was a better preacher than he was a blacksmith. A man may be a better prose-writer than a poet or a better thinker than a doer. A teacher's training may be better than her experience, or she may be a better teacher of Latin than of sewing.

Indeed, it is continually necessary in daily life to compare things which, so far as they are measurable at present or conceivably measurable at some future time, must be expressed originally in units of different kinds—such units as dollars, pounds, years, words spelled, problems solved, and errors committed. Only by converting the number of these original units possessed by a person or thing into some unit of variability such as the median deviation (or the standard deviation or the average deviation) do we strike common ground. This assumes that we have collected many measurements like the particular ones which we are comparing and that from a distribution of them we have derived, either statistically or by the less formal process of conceptualization, a type or average on the one hand and a measure of variation from the type on the other hand. Thus a given child is provably a better reader than speller, because knowing what average reading and spelling are and the variability of children about these averages, we find that he stands better with respect to the reading average and in terms of the reading variability, than he does in respect to the spelling average and in terms of the spelling variability.

When we reflect as to what is meant by saying that a man is a better preacher than he is a blacksmith, we realize that even here we have a rough though by no means easily communicable notion of an average preacher and an average blacksmith, and that *as preachers run* (i.e., according to their variability) the man in question is judged to exceed the preacher average by more than *as blacksmiths run* he exceeds the blacksmith average.

Thus, the two factors both in statistical and conceptual thinking are *type* and *variation*. Indeed in all our communication of thought which exceeds the Scriptural "Yea, yea" and "Nay, nay" these two facts are implicitly present.

It is clear, therefore, that in the formal processes of statistical method we cannot properly dissociate these two complimentary data. In particular, we cannot rely on averages alone. Many just criticisms of statistical statements have been incurred be-

cause nothing but averages have been reported and used. An average tells nothing at all about the extent to which items are grouped about it. Hence, in judging about a given magnitude it merely constitutes a point of departure.

For example, two teachers each receive a salary of $1,200. One is in a city in which the average salary is $1,000 and in which no salary falls below $800 nor exceeds $1,200. The other is in a city in which, although the average is likewise $1,000, the range is from $500 to $2,000. If we only know the average, all we can say is that each teacher is getting $200 more than the average. With only this amount of information the conditions look alike. In reality the teacher in the first city is receiving a high salary—the maximum—while the teacher in the second city is receiving relatively a much lower one.

The essential basis for the really significant statement of the amount of departure from the average is the natural, common unit of variability. Moreover, this variability is just as truly a norm or standard as the average which is so generally and so exclusively regarded as such. In fifth-grade arithmetic, for example, it is as valuable to know the typical variability as it is to know the typical central tendency. A teacher is as truly concerned to know how much, in point of intelligence, pupils differ from a standard variability as she is to know how much they differ from the average. Indeed, it is perhaps reasonable to believe that she is more interested and more affected in her work by the diversity of talent than she is by its general level.

Since, therefore, in language and in statistics a measure of variability is essential, since it forms the basis of our judgments and enters into all our thinking, since it is the natural unit in which all magnitudes may be expressed and in terms of which they may be compared, since it is as truly a norm or standard as the average itself, and in short since for all these reasons it is the indispensable supplement of the average in characterising both groups and individuals, we urge that in educational reporting it be more systematically presented, that in reaching decisions it be more generally regarded, and that in teaching and administrative procedures it to be given its rightful place.

We feel sure that if research workers more generally report measures of variability, and if they more fully interpret them and show their significance, these measures will enter more fully

Into the consciousness of school people to refine their thinking and to make their judgments more discriminating. Here, as everywhere, the real test of the efficacy of research work is to be found not merely nor chiefly in the extent to which it satisfies technical requirements but in the extent to which it meets the practical needs of the hour.

B. R. B.

"SELLING" EDUCATION

Education is a great enterprise. Locally it may appear small. The neighborhood schoolhouse, even though it may not quite be of the "little red" variety, is frequently unprepossessing. Yet altogether apart from the emotional notions of the uplifter and of the demagogue when he is after the support of teachers, education as an institution is big in a material sense.

One-fifth of the inhabitants of the country are attending school, and practically everybody else has either attended or will attend. It is the only establishment save government itself which assays to minister to all the people. Nearly a million persons are engaged in conducting its affairs. Not infrequently one-third to one-half of the proceeds of taxation are devoted to its support. It is a business whose offices are everywhere—not only in every populous city, but in every village and hamlet. Its holdings, in mere acreage, constitute a principality. Its output is trained boys and girls—a product upon which no value is set, because it is priceless.

Big, however, as education is, its size is not appreciated by the people. It is so near at hand and so modest in its local setting that the general public fails to realize that it is, in fact, near *everybody's* hand and that it has as many local settings as there are localities. The schoolhouse and the school taxes, the school teacher and the school children are so much a part of our lives that we do not realize that all these things in their present form are distinctly new and peculiarly American. We fail to understand that substantially every square mile from coast to coast and from Canada to Mexico is part of a school district.

The people are entirely too prone to accept education as a matter of course. We have it where we are now living and we shall have it if we move elsewhere. We have it today, and we shall have it tomorrow. Why take thought, then, for this institution which is always and everywhere present?

To be sure, some of us know that education is by no means a matter of course. To us it is a complex and difficult organization—interesting, ever new and vital. Its problems are of far-reaching importance and their name is legion.

But we must bring this conviction to the people. We cannot sustain education by our own efforts. The most fundamental necessity is public support. None of the solutions of our problems can be put into effect without money.

With this in mind, and with the determination to keep faith with our cause and with ourselves, we must exalt education as an institution in the minds of the people. In the lingo of business—we must "sell" education to the people.

Nor should this prove to be difficult. We have a great cause. In it the people, though they may appear apathetic, have an abiding faith. Even a moderate degree of effort may be expected to produce large results.

Contrast, if you will, the task of selling education with the task of selling insurance or automobiles or patent medicine or chewing gum. Note the skill of the advertiser in creating among a large number of people wants more or less unnecessary and hitherto entirely unsuspected. Observe the adroit manner in which the salesman displays his wares. The selling campaign is conducted with a high order of business ability. We cannot withhold our admiration, although we may feel that this ability is often devoted to unworthy ends.

Education, on the other hand, is already half sold. To sell it completely is perhaps the easiest of selling enterprises. A little more than a year ago, we attended a conference on the educational crisis called at Washington by the Commissioner of Education. This conference was attended by several newspaper men who were more or less interested in education. It was their unanimous opinion that to arouse the people to the importance of education was, as one of them said, "a lead-pipe cinch." This particular speaker went on to show how, unlike other enterprises, education could count on a fundamental interest from the beginning; how, therefore, interest would not have to be created; and how wants would not have to be manufactured. Nevertheless he showed that the material which was being furnished to the press on the subject of education was ordinarily so poorly prepared as to be of doubtful value.

Believing, as we do, that almost if not quite the most vital need for the immediate future is a true sense of the value and dignity of education as a democratic institution, we urge the necessity of selling education throughout the country. The method is clearly that of publicity. If the interest of the people is real though latent, we should arouse it by legitimate publicity methods until it becomes a compelling force. There is nothing more important today than that this work be done, and that it be well done. There is no lack of material; and if this material is properly presented, there will be no lack of available channels of publicity. Every group of teachers should have its publicity committee—not a committee for securing better salaries, but a committee for placing before the people the purposes, the problems, and the achievements of the public schools. Not only elementary and secondary schools but every form of higher education should be included. Every newspaper within the city, county, or state, as the case may be, should be reached with simple, interesting, and continuous copy. Alliance should be sought with the churches, with Rotary Clubs and similar organizations, with Chambers of Commerce, with agricultural societies, with parent-teacher associations, and with labor organizations. In fact wherever people are brought together in such a way as to exert an influence upon the public or a portion of it, there the publicity campaign should seek to be effective.

Nor should this publicity movement be, as it has been in the past, a sporadic movement, hastily gotten up, directed toward a narrow objective, and abandoned when a decision concerning the objective has been reached. It should be a continuing policy. Its organization should be essentially that of taking the people into the confidence of those who are carrying forward the work of training children and young people.

In this enterprise it is clear that information, accurate, reliable, and significant, must be available. Moreover, it must be presented attractively. Workers in educational research are precisely the persons who have, or should have, this sort of information. At present, too much of it is being kept in our files or communicated to each other without becoming widely known. We realize, of course, that our researches must eventuate in some form of record or report. Let us realize that in addition to this we have a far greater task to perform; namely, that of supplying the data, as well as of participating in the organization, for a campaign, continuous, far-reaching, and disinterested in *selling education to the people.*

B. R. B.

Reviews and Abstracts

ARPS, GEORGE F. *Work with knowledge of results versus work without knowledge of results.* (Psychological Review Monographs, whole no. 125, Volume 28, no. 3.) Princeton, New Jersey: The Psychological Review Company, 1920. 41 pp.

The monograph reports a study with the ergograph—an instrument which lends itself well to the investigation of the progress and effects of fatigue. Roughly, such an instrument consists of an arm rest into which the subject's arm is firmly strapped; and the work consists in either raising a weight or pulling at a spring by repeated flexing of a finger, to which a cord is attached. The purpose of the apparatus is to exercise repeatedly a single very restricted set of muscles. The arm rest prevents pulling at the cord by moving the whole arm or otherwise than by means of the flexor muscles of the particular finger experimented with. The finger works against a known resistance (in most of the present experiment the load was four kilograms); and in other ways the situation is controlled so that it is possible to study with a considerable degree of exactness the progress of fatigue and factors conditioning efficiency, as these factors appear in the repeated functioning of a single relatively isolated part of the neuromuscular apparatus. In the experiment here reported the right forefinger was used; and it was flexed once each second, as timed by a metronome. After every ten flexions there was a rest period, varying from 0 to 10 seconds. The experiments were made every other day; and each day the work was continued until the finger was "exhausted" —that is, failed twice in succession to move the weight. The very important and very practical question was whether there was any difference between the amount of work done when the subject knew his previous records and could observe his record-of-the-day as he made it, and the amount of work done when the subject knew nothing of his record. Three subjects in all took part in the experiment; and one of these subjects continued in the work three years, thus becoming very thoroughly habituated and trained to the procedure.

Several very interesting results appeared. In the first place (the major conclusion) the amount of work done averaged distinctly greater with knowledge of results than without knowledge of results. It is pointed out that "Will power as conventionally regarded is inadequate to explain the efficiency differences"; the subjects throughout were instructed to make maximal effort. Rather, knowledge of results operated in some fundamental way to be "provocative of greater functional changes in the central nervous system." Work without knowledge was reported by the subjects as very deadening, incapable of being sustained, lacking in vitality.

In the second place, however, in a series without knowledge of results, run *after* a series with knowledge of results, it was found that imaginal elements tended to sustain the activity in somewhat the same way that actual knowledge of current results had done. Under these circumstances the subject imagined the record of each individ-

ual lift, comparing the strokes of the recording stylus in successive lifts, and so compensated, in a way, for actual observation of results.

In the third place, when work has progressed close to the point of exhaustion "a curious phenomenon of sudden recovery appears." In other words, when the movements have almost ceased there is a sudden recovery to efficiency almost as great as at the beginning. Indeed, this sort of recovery may take place a number of times.

These experiments (which may seem to the average school man, at first thought, to have little other than a theoretical importance for professional psychologists) really have a very interesting bearing on many current educational problems. The findings are surely one more excellent argument for systematic practice exercises in the various school subjects with individual records of progress accessible to the children or (perhaps) kept by them. Apparently work "with knowledge of results" not only gives better motivation but operates in some more fundamental way to increase efficiency. The sudden recoveries are also of no little practical interest. Evidently even in very simple forms of work there is a variety of compensatory mechanisms, the situation is highly complex, and the nervous system is protected in a number of ways from actual complete exhaustion of the apparatus involved in any one activity.

The study is based upon a mass of data which have been analyzed in most painstaking fashion. It is typical of the best type of laboratory work—a type of work of which there has been all too little recently, as a result of distraction of effort into the more spectacular activities of "testing." It is to be regretted that there is not more "transfer" of such laboratory findings into current educational thought. For the most part such psychological investigations as this seem to escape the notice of the educationalists.

S. L. Pressey
The Ohio State University

Wilson, G. M., and Hoke, K. J. *How to Measure.* New York: The Macmillan Co., 1920. 285 pp.

This is the latest book in the field of educational measurement which has come to the writer's attention. The volume contains twelve chapters as follows: The New Attitude Toward Measurement; The Measurement of Spelling; The Measurement of Handwriting; The Measurement of Arithmetic; The Measurement of Reading; The Measurement of English Composition; The Measurement of Drawing; The Measurement of Other Grade Subjects; The Measurement of High School Subjects; The Measurement of General Intelligence; Statistical Terms and Methods; The Teachers Use of Scales and Standardized Tests.

In each of the chapters devoted to a single subject a number of the tests and scales appear complete with directions for use. Standards of achievement or grade norms are given and numerous distribution tables and graphs illustrate the treatment of data. Each chapter except the first closes with a bibliography on the field discussed in the chapter. The volume is written in an easy running style which can probably be read for the most part by the average teacher with little difficulty.

The authors make two statements in the preface by which it seems fair to judge the contents of the book. 1. "The present volume is dominated by two main ideas, first, that the work in measurement should be handled more and more by the individual classroom teacher; and second, that the chief purpose to be served by standard tests is the diagnosis of pupil ability and pupil difficulties." From the first clause of this

pronouncement one would expect the book to be devoted to the task of teaching the teacher the use of educational measurements. There is much evidence that the authors endeavored to do this but it is also clear either that they could not resist the temptation to include data which were at hand though not needed for this purpose or that they had other purposes in mind. For example much of the material in tables 4, 11, 14, 17, 24, 25, and 41 would not have been needed if the main ideas above stated had completely controlled. These are data of interest to supervisors rather than to classroom teachers.

Having declared that "the chief purpose to be served by standard tests is the diagnosis of pupil ability and pupil difficulties," one would expect the book not only to show the teacher how to locate the children having difficulty and the type of difficulty of each, but also to be full of specific helps for remedial work. Considerable attention is given to teaching the teacher how to locate the children, and some little effort is made to show her how to diagnose their difficulties; but little aid is given on the most serious problem—that of remedial treatment. In fact, the authors state specifically more than once that "it is not the province of the present work to discuss methods in any extended way, but merely to show the use of standard tests." (The quotation is from page 71, but the same idea is expressed on pages 21 and 43). Has not the day passed when experts in any field may consider their task done when they have pointed out ills or the means of discovering ills? It is true that the headings "Remedial Instruction and Remedial Measures" occur four or five times and that some suggestions are given under the title "Using the Results"; yet it can be clearly shown that the major emphasis is upon the form of the various tests and the standards derived.

2. The second prefatory statement by which one should examine the book is: "The purpose of this volume is not a critical evaluation of all the available tests on different subjects, but a treatment of those tests which on account of their use, purpose and adaptability have been found to be most serviceable to the classroom teacher." Those thoroughly familiar with the field of tests and measurements will doubtless wonder at the omission of Greene's Organization, Minnick's Geometry, and others which have been on the market for some time. If these are ruled out under the criteria above, one may ask why Rice's Spelling Test is given in full and why considerable space is devoted to various other tests of doubtful value.

E. J. Ashbaugh

Ohio State University

Proctor, William M. *Psychological tests and the guidance of high school pupils.* (Journal of Educational Research Monographs, No. 1.) Bloomington, Illinois: Public School Publishing Company, 1921. 70 pp.

This monograph is based upon an extended study of the relation between intelligence scores of high-school pupils and their educational success, as indicated by school marks, teachers' ratings, elimination, etc.

The study is of broad scope and deals with the important educational problems involved, rather than with the technic of intelligence testing. While the method used and the conclusions drawn are admittedly tentative, the monograph contains a large amount of material which will prove extremely helpful to high-school teachers and principals. The investigation extended over several years, thus giving sufficient time to check up the test results by records of the later success of the pupils. The

work was begun in 1916-1917 when the Stanford-Binet scale was given to 137 pupils in the Palo Alto High School. These pupils have now been followed for five years.

In 1917–1918 and in 1918–1919, 955 pupils in five high-schools were given the Army mental tests. The results of these have also been followed up and correlated with school success and elimination. Correlations of both the Stanford-Binet and the Army mental tests with school marks were in all cases high enough to show the tests to be of considerable value for educational guidance. Pupils who have low scores tend to predominate among those who drop out. Average scores of pupils enrolled in certain types of subjects differ greatly from those enrolled in other subjects. Score limits are indicated for the tests used, below which pupils are very likely to fail in Latin, algebra, etc. As would be expected, school marks in the semi-vocational subjects show low correlation with intelligence scores. The conclusion is obvious. There are plainly a great many children who, because of lack of ability to think in abstract terms, cannot succeed in mathematics, Latin, or English literature, but who can do fairly creditable work in household science, shop work, etc.

The reviewer believes the author's findings along the line indicated above to be extremely significant. Clearly it is going to be necessary to work out the chances of intelligence of various grades succeeding in the different kinds of school work. Until we have such information, educational guidance will be out of the question.

Another important problem is that of shaping the instruction in a given subject to the abilities of the children. Since the range of ability in any class is likely to be very wide, schools are compelled to consider the desirability of sectioning their classes on the basis of intelligence. This practice is becoming common in the grades, and is no less important in the high-school.

The author's chapters on the use of psychological tests in educational guidance and vocational guidance are based upon considerable experience in carrying on this kind of work both with high-school pupils and with Federal Board students in the university. The data presented show in a very convincing way that the intelligence score is always worth considering in advising any pupil regarding his future educational or vocational work.

Extensive data are presented on the university success of the high-school pupils studied. It was found, for example, that school marks of first-year university students at Stanford correlated to the extent of 0.46 with Stanford-Binet tests of these pupils given five years previously. This is one of the longest range comparisons known to the reviewer, and shows conclusively that intelligence test scores have considerable permanent value.

The monograph is a careful study of a problem which is certain to occupy the attention both of psychologists and of school men to a far greater extent than has been the case during the past. The importance of the problem is being rapidly emphasized by the passage of laws in many states designed to retain all children under the jurisdiction of the public school until the age of eighteen. The public schools of the future must deal with the entire child problem. They can no longer set arbitrary standards and eliminate the pupils who are unable to work up to these standards. Intelligence tests are necessary to help the school in making this adjustment.

LEWIS M. TERMAN

Stanford University, California

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors.

Adapting American Tests for Use in China

William J. Lacy, one of our correspondents in China, writes interestingly on the adaptation which is being made in China of the tests we are using here in America. Mr. Lacy is the executive secretary of the Conference Board of Education, Methodist Episcopal Church, Yenping, Fukien, China. We reproduce his letter in full:

Your letter of April 8th has been waiting my return from North China where I have been spending two months in famine relief work. I thank you very much for your invitation for me to contribute some brief report of the educational research work in which I have been engaged for the JOURNAL OF EDUCATIONAL RESEARCH, and shall hope to send you something within a few months.

At present I am at work on an adaptation of Daniel Starch's Arithmetic Reasoning Test which has been put into Chinese and is being used in our higher primary schools here in this province. The higher primary school is the top half of our American Grade School so that the tests are being used in our grades corresponding to the American 5th to 8th grades.

My initial investigation was carried on a year ago this spring at which time I was able to establish score values for the various problems which I had put into Chinese in a revised form. This year I have had the tests given in about thirty different schools and when all the returns are in and I have tabulated the results I hope to have a fairly well established standard score for each grade. It has been necessary, of course, in adapting these tests to make certain changes in subject matter as our American money values have to be rather radically changed because of the complicated monetary system in use here in China. Just to show you some of the difficulties in making an arithmetic test where money problems are used—and we certainly want to use the money problems because money plays the greatest part in the thought of the Chinese—I will give you the different steps in the changing of a dollar.

Money is spoken of in this part of the country as big money and small money. Big money consists of the dollar silver piece as standard. A dollar silver piece will change into eleven dimes and from one to four pennies, owing to the prevailing rate of exchange on a particular day. Then each dime changes into from eleven to twelve—usually twelve—pennies, so that a big dollar is worth anywhere from 130 to 135 or sometimes 140 pennies. Now, when you go to put any of our American arithmetic problems which deal with money into the hands of your Chinese, you have a problem that is not easy to solve. We have therefore more or less used the small money standard which consists of ten dimes to the dollar, but your dimes are again changed into more than ten pennies, so that in any such problems you either have to deal entirely with dimes and half dimes or else state the rate of exchange for pennies that is to be used in solving the problems. This, of course, puts another element into the solution of the problems and vitiates the results when it comes to any comparison of the same problems used in America.

I found in my initial investigation last year that the median score made by the Chinese pupils in the adapted tests which I had made was a little bit higher for the

same grade than that for the Starch Test at home. On the other hand in any test where the time element is used the Chinese students are much slower than American students.

What I have written is merely to give you a brief suggestion of what I am working on at present. I hope some time this summer or in the early fall to be able to complete the compiling of my results and get out a full report of the work that I have done with these tests. I shall then send you a copy of this report, and you can use such parts of it as you choose.

Use of Intelligence Tests in the Mission Schools of China

Mr. Fred P. Beach of the Department of Education of the Fukien Christian University of Foochow, China, gives us the most complete statement we have yet received as to the function which intelligence testing has or may have in the peculiar educational situation confronting educators in China. From what he says it is entirely clear that this function is by no means less than it is in America. Mr. Beach, like Mr. Lacy whom we quote above, is an example of the highly trained expert whom the missionary service is now enlisting. These men are evidently engaged in a momentous undertaking—an undertaking to which their enthusiasm and ability nobly corresponds.

We are sure that our readers will find Mr. Beach's letter of unusual interest. It follows:

The real situation as far as tests are concerned is about as follows: They will be, when really ready and well-tried-out, an invaluable aid to us not only in our school work but in all of our missionary effort. Some eight years ago Dr. Paul Munroe came through here and I had a chance to visit with him a little. One of his remarks has stuck in the minds of some of us, namely: that the missions cannot hope to succeed in their conversion of a nation by the use of unintelligent material. Or as I paraphrased it "We cannot hope to make St. Pauls out of defectives, delinquents and dependents." Now it is the truth that the early mission had to take anything that came to it; and amongst the lot was a great deal of the type of the dying-out-family. There is nothing so strong in China as the Family. There is little individuality as we know it, and the successful man in China must support his relatives. The richer he becomes the more relatives he takes on. You can see that a Chinese without a family is in a pretty bad social and economic way. In their straits they have turned, many of them, to the church. Now the problem put up to us has been the education of their children. To a large extent the missions have been compelled to adopt the tactics of running very expensive schools and of subsidizing through these schools, those of its own Christian children who were bright enough. This is not the ideal way—not very democratic, perhaps—but as long as the mission schools continue to teach the coveted "English" better than the government schools so long we will continue on this tack. Now you can see that on this basis of operation a great many dollars could be saved to the missions if it were possible to give intelligence tests beforehand to these students and not have to wait for a year's school work to show up the poor students—and even then perhaps have to yield to importunity and give the child a second year. On the other hand, it will be as useful to us as to you to be able to pick out the extra bright and give them a better course.

There are probably no harder working students in the world than our Chinese. Nevertheless, we are facing in our schools the same problem of "not allowing your studies to interfere with your college course," and it is a powerful stimulus to be able to tell the students in chapel that intelligence tests do *not* test moral character or fairness, and that it is perfectly possible for a man to have high intelligence (I. Q.) and yet fail at the examinations, as a man "cannot remember what he has not seen."

Now as to technique. To tell the truth we are mere amateurs in the beginning stages of experimentation. Knowing Terman's book the best, we have based our work mostly on his. It is obviously impossible for us to pay fancy prices for printed tests by the hundred or thousand, and one wonders if they are any the better for being put

up that way. We shall have to develop some cheaper way to give group tests. Of course there is no copyright law and only our consciences will prevent us from pirating anything we need. But on the other hand *our main problem* is not the finding of American material and translating it into Chinese. The Translation problem is as easy as can be. Any University or Normal in China will have at hand some returned student who can translate perfectly. The *real question is the validity for Chinese society of the American or European Tests.* This is our experimental task for years to come. Inasmuch as ones "system of knowledge" (Pillsbury) is only built up by experience in this world as it is, these tests obviously must test the ability as it has been used in its social environment, and the good test is evidently that which is new to the person at the age tested and yet which depends on a certain definite experience and growth. In other words, your relativity again.

It is not difficult to change the Standford Binet tests so that instead of "the engineer going faster with a heavier train" it becomes a ricksha coolie with more people in his ricksha. But I take it that our task is to test thousands of children all over China till we have coördinated our work and gotten a pretty reliable idea of what can be expected of Chinese youth and children in such a social environment as they grow up in.

Meantime we have more than enough opportunity, as the tests are desired by all our schools. Our university will give a course in the Psychology Department next semester in order chiefly to get a set of students at work on the tests among the school children. Meantime we shall watch you fight out the validity of the underlying principles. I believe the tests America is turning out are useful even now, not as an absolute guide to whether a student shall be dropped or not, but as an additional help confirming in a general way the poor work of the student. I think we can safely use them in addition to the entrance examinations for those entering the High School or College.

Inasmuch as our Educational Association is striving now to supplant the once popular but now largely disliked "Uniform Examinations" and is trying to put "Standard Educational Measurements" in their place, it is obvious that any fair comparison of country schools as against city schools should be able to say what the median I. Q.'s of the schools compared are. The country schools try to do as much as possible for slow students whereas the city mission schools are rather relentless in dropping students who get a little behind.

In conclusion, I have no doubt that the tests may develop some national strengths and some weaknesses. Still I am confident that they will bear out the accepted theory that the Chinese are one of the great dominant races of the earth. The old examinations were in effect a test of intelligence; not a full rounded one, and certainly not a moral test at all. But for the most part it is the children of those who succeeded in the old tests that produce the greatest percentage of good students in our schools. Amongst our own church children there is great variation. But fortunately among them are a few of high ability, thanks probably to Mendelian proportions; and from these the church will get some leaders and it has already. Now, however, we should like to be able to know them when we meet them.

You may be interested in studying over the following vocabulary problem. In a country where ideographs are used and where verbal memorizing is still largely used, is a vocabulary test from the dictionary an intelligence test or is it a "measurement of classroom products?" This is a matter of immediate concern and is being worked out in several places in China. Notably Pekin and Nanking. We are doing some work here and will try some more but haven't gotten very far yet. To quote Mr. Hocking, "One doesn't need to apologize for not being perfect."

Before closing up his letter Mr. Beach adds this postscript:

Since writing the above I have given Terman's 100 word English vocabulary test to 30 freshmen in our University. Rather a wild chance perhaps, but I wanted to see what they would do with it and how they would compare with American children. I found that the best knew 51 words while they ranged down to 32, 65 being the number for "Average adult" and 75 for "Superior adult" in America. This, of course, raises the same question, "How much is this a test of classroom product and how much is it a test of intelligence?" It would be necessary to try this on all freshmen in China to

get any relative information. But it does indicate that there is a new angle of approach to the teaching of English in the Middle Schools here and that such a vocabulary test may well be used as a classroom standard for the students. I see that I must go on and make the same test on upperclassmen. The freshmen above mentioned were all in the neighborhood of 20 years of age.

On Burgess' "The Education of Teachers"

The following comment by Charles Carroll of the office of the Commissioner of Public Schools of Rhode Island together with a reply by W. Randolph Burgess came to hand during the summer. Mr. Carroll speaks first:

The JOURNAL OF EDUCATIONAL RESEARCH for March, 1921, printed "The Education of Teachers in Fourteen States," by W. Randolph Burgess. May I not be permitted to comment briefly upon the methods pursued and the conclusions drawn by Mr. Burgess, in the interest of greater accuracy in the use of educational statistics?

1. On page 161 the author quotes statistics of the educational qualifications of Massachusetts teachers, without indicating clearly that the statistics do not include all teachers employed in Massachusetts. Generally the Massachusetts statistics on the educational qualifications of teachers have been for full-time teachers only, and have omitted Boston. The omission of teachers employed for parts of years only, or otherwise irregularly, should have a tendency to produce averages more favorable than otherwise, in accord with a fair presumption that those well qualified probably will have more regular or full-time employment than those not so well prepared for teaching.

2. In the table on page 166 the writer has taken for Massachusetts the number of full-time teachers employed outside Boston plus an estimate for Boston based on the advance report of the Commissioner of Education for 1920. The table on page 161 indicates a significant improvement for 1920 over 1914 in Massachusetts; perhaps an estimate for 1918 based on figures for 1920 should be discounted by at least 2½ per cent, or one sixth of the almost 15 percent increase between 1914 and 1920.

3. In the same table the estimated figures for Massachusetts full-time teachers are compared with the Rhode Island figures for all teachers, an objectional procedure for the reasons indicated in paragraph 1.

4. The author's interpretation of the diagram on page 167 appears to be incongruous if he is really attempting to find a fair index of preparation for teaching, because he fails to distinguish "mere schooling" from "professional training." Massachusetts is accorded first rank because of having 17 percent of college graduates amongst its teachers. Rhode Island's unsurpassed 70 percent of normal graduates is ignored or disregarded. In other words, the author appears to have accepted the hypothesis that preparation for teaching is to be determined by years of schooling. If the hypothesis were correct, a medical doctor with four years of college training plus four years of university instruction in a medical school, should be four times as good a teacher as a graduate of a two-year normal school. No one should question seriously the superior preparation of the doctor for the profession of medicine; but it is no more logical to hold that the doctor's education has fitted him for teaching than to assert that two years of study in a normal school is one-fourth of the preparation for practicing medicine.

5. In computing index numbers, as shown on page 168, the author has ignored the fact that in Rhode Island the normal-school course at Rhode Island College of Education has been 2½ years. Were credit given accordingly, the index numbers for Rhode Island would be 2.352, and Rhode Island, even on the count of all teachers as contrasted with only full-time teachers as in Massachusetts, and on accurate figures instead of estimates for Massachusetts, would lead Massachusetts and all other states.

6. In Rhode Island graduation from college is not recognized as adequate preparation for teaching. For many years the standard requirement for professional certification of college graduates has stipulated six semester courses in the science and art of education supplementing the requirements for the bachelor's degree. Because of this standard minimum requirement, and because of the relatively large number of

college graduates teaching in Rhode Island who have pursued additional graduate courses in education on state scholarships leading to masters' and doctors' degrees, it would be reasonable to rate the Rhode Island teacher who is a college graduate as having had at least an average of 4½ years of preparation beyond graduation from high school. Were credit given for this also, Rhode Island's index number would be 2.424.

Mr. Burgess may not have been familiar with either the length of the course in Rhode Island College of Education or the certificate requirement for college graduates. In correspondence preceding publication of his article, however, his attention was directed to the fact that percents for Rhode Island based upon a full count of teachers probably would not insure full credit for the State.

Charles Carroll

Mr. Carroll's comment was sent to Mr. Burgess who submitted the following statement by way of rejoinder:

A complete answer to Mr. Carroll's letter would involve a detailed discussion of the records and computations back of my article in the March number of the JOURNAL OF EDUCATIONAL RESEARCH which would not be interesting to most of the readers of this publication, and would tend to be contentious. The reader who is interested in doing so may form his own conclusions by carefully reviewing my article in the light of Mr. Carroll's letter. The questions which he raises had all been carefully considered before the article was completed, and concerning most of them specific comment is made in the article. The state reports from which figures were taken are accessible.

I cannot help feeling, however, that Mr. Carroll's real difficulty does not arise from a difference of opinion as to specific details of my article. I think it rather arises from a difference of view point concerning the major purpose of the article. In the past few years statistics have become more practically useful than ever before. The reason lies largely in the wider development of the method of sampling, or representative returns. We are now able to measure price changes by a price index number which is based, not on a tabulation of all prices there are, but on a selected few. We are learning to measure human intelligence, not by tabulating all of the acts of an individual, but by studying sample behavior. In my article I wished to suggest a possible method by which we might secure a widely practicable measurement of teacher preparation. Completely exhaustive measurement of preparation is not possible. It is possible, however, to measure the situation by the method of sampling. This statistical method has always been open to the criticism of the person who calls attention to the specific cases which the method does not measure, but those who have experimented carefully with the method know that it, nevertheless, may give substantial truths. It is clear that some method of measuring teacher preparation is greatly needed. I think that the method suggested in my article would yield, with some further refinement in the collection of data, exceedingly valuable results. Even with the imperfect data now at hand the method yields results which are significant.

W. Randolph Burgess

On the Assumption That Errors of Estimate Are Equal in Narrow and Wide Ranges

Professor Kelley's statistical article in our May, 1921, number evidently evoked some interest—a mild and entirely decorous interest as would become those who are statistically minded. We received a comment on this article (or shall we call it a criticism?) from Doctor Karl J. Holsinger of the University of Chicago. He has located a really vulnerable spot in Professor Kelley's article, namely the assumption that errors of estimate are the same over wide and narrow ranges of talent.

This assumption, it will be recalled, occurred in the course of an argument to the effect that unless we know something about the spread of ability in the subject and relative series we cannot interpret the correlation coefficient. Professor Kelley's

specific statement here was that a correlation coefficient of +0.90 for children of the fourth to the eighth grades taken as one group might mean no closer relationship between the traits in question than a coefficient of +0.40 for children of the fourth grade only. Independently of the appropriateness of the specific criticism which Doctor Holzinger makes, this point is exceedingly well taken. It is worth repeating we cannot know the real meaning of a correlation coefficient until we also know something of the range of the paired measures which enter into it. To this we judge that Doctor Holzinger, as well as every other thinker along statistical lines, will agree. Nevertheless, his position is well taken. He writes:

I have been much interested in an article by Professor Kelley on The Reliability of Test Scores, appearing in the May number of your JOURNAL. There are some points in the article that appear to be at least debatable—in particular the basic assumption that "the two errors of estimate are the same" on narrow and wide ranges (bottom of page 377). It would seem to be equally plausible to assume that the above errors are proportional to the corresponding "spread of talent" as given by the "true" sigma's. Under the second assumption, of course, r and R will be equal if the factor of proportionality is the same for errors and true sigmas. In other words, if we assume $\Sigma_{1.t} = k\sigma_{1.t}$, we have

$$\frac{k\sigma_1}{\Sigma_1} = \sqrt{\frac{1-R}{1-r}}$$

and for formula (2)

$$\frac{k\sigma_t}{\Sigma_t} = \sqrt{\frac{r(1-R)}{R(1-r)}}$$

Whence, if $\Sigma_t = k\sigma_t$, $r = R$.

Of course, all this is self-evident; but it shows the violence of Professor Kelley's hypothesis (cf. $r = .4$, $R = .914$, page 374). Spurious correlation due to age factor and heterogeneity will increase correlation for a large group such as Professor Kelley describes, but it is doubtful whether his formula properly accounts for such increase.

I am writing to ask you if you will give your readers the benefit of some editorial comment on some of these points, if you think it worth while. This matter of reliability is so important that a critical review such as Dr. Kelley's is most helpful, and might well be followed up by such comment as I suggest. I wish to express my high regard for Professor Kelley's work in general and in the present article. It is because of this that I am particularly anxious that we should agree on his methods.

Instead of commenting editorially on this matter, as Doctor Holzinger suggests, we have preferred to submit his criticism to Professor Kelley for his comment. This procedure commended itself to us for two reasons. First it meant less work for us and second the result would be better. The reader may note, if he pleases, the order in which we state these considerations.

Professor Kelley replied almost immediately and we at once felt justified in resisting the temptation to comment editorially. In particular, we should not have added "homoscedastic" to the vocabularies of some of our readers.

We wonder how we ever managed to get along without this handy term. And its derivative, homoscedasticity, is even better. Do not these words fill a long-felt want?

Professor Kelley's letter follows:

In regard to the point raised by Mr. Holzinger let me say that the existence, or non-existence, of the equation $\Sigma_{1.t} = \sigma_{1.t}$ is an experimental matter and not a mathematical necessity, holding for all kinds of scatter diagrams or correlation surfaces. It is intimately connected with homoscedasticity, or equal variability of arrays. Galton

run into the same problem and found that the arrays with which he dealt were substantially equally variable. In connection with a heredity problem he wrote:[1]

"I was certainly astonished to find the variability of the produce of the little seeds to be equal to that of the big ones; but so it was and I thankfully accept the fact, for had it been otherwise I cannot imagine from theoretical considerations how the typical problem could be solved." Galton's work led to the discovery of the normal correlation surface and in this arrays are equally, not proportionately, variable. I can think of no more sound assumption to make than that the correlation surface involving a single test score and a "true" test score constitutes a homoscedastic system[2]; but I will be the first to grant that it is an assumption, to be checked against experimental findings wherever possible. Should it not prove so we must regretfully accept the fact and exercise greater ingenuity than Galton imagined himself possessed of, to discover the lines of solution of the problem.

The old problem, reiterated by Mr. Holzinger, involves at least two things (a) the nature of the variability of successive arrays, and (b) the equivalence of the units in a test or scale. Until these are experimentally determined for a given test, I think we shall make fewer errors in interpretation if we assume that arrays are homoscedastic (equally variable) and that the units of measurement throughout the test are equal.

In closing, I would say that with reference to scales of the Binet type, which assume equivalence of successive age intervals, I think we already have abundant evidence to refute assumption (b). Accordingly, I do not believe the equation $\Sigma_{1.t} = \sigma_{1.t}$ holds for narrow and wide ranges in the case of the Binet scale. Not having two comparable forms of the Binet scale this matter is somewhat difficult to test, but a way for doing so can surely be devised.

Should the preceding observations tend to enlighten rather than confuse the readers of the JOURNAL I should be glad to have you print them in connection with Holzinger's criticism.

National Association of Directors of Educational Research

(E. J. ASHBAUGH, *Secretary and Editor*)

Perhaps nothing that can be reported in this issue of the Journal will be of greater interest to our membership than a statement concerning the meeting of the Executive Committee last month. Because of the removal of Dr. Buckingham from the University of Illinois to Ohio State University as noted in last month's items, certain questions arose as to the relation of the Association to its official organ. Final action of the Association cannot be given at this time, but it should be said that the Executive Committee after a thorough analysis of the situation agreed unanimously to stand for a journal of educational research whose board of editors and editorial policy should be guided by the Association.

A large portion of the time of the meeting of the committee was given to the tentative formation of the program for our next annual meeting. President Rugg informed us that he had been notified by the secretary of the Department of Superintendence that our association is one of the TWELVE organizations invited to cooperate with the Superintendents at their Chicago meeting. This was encouraging news, showing that the great organization, which declared independence in control of its own meeting last winter, recognized the interest of its members in the work of our association and the real value of our programs. Certainly this year to an even greater degree than ever before, we may expect a hearing. We shall undoubtedly have an opportunity to impress the educational leaders with the importance of our work. The biggest and best program ever is the ambition of the president upon whom

falls the burden of securing the participants. We may have full confidence that he will realize his ambition.

Remember the place—CHICAGO.

The time—Last week in February.

The purpose—Selling the work of the National Association of Directors of Educational Research to all good school people.

You—Make your plans and reservations now.

The 1918–1920 Biennial Report of the State Department of Public Instruction of Wisconsin entitled "Educational Progress in Wisconsin" and edited by Mrs. C. W. Flemming has just come to your secretary. Mrs. Flemming, who is leaving the department to spend this year in Teachers College, Columbia University, leaves behind her a well-written report of work done and a valuable list of suggestions for the future work of the bureau of educational measurements in the department. Only a summary of the summary can be included here, but it is hoped that the department was able to send a copy to each of our members. The first paragraph will be quoted since it can scarcely be condensed. "The effort of the first two years of state supervision by means of educational measurement (1916–1918) was (1) to convince the school men of the state of the need of experimental study of school problems; (2) to familiarize them with the technic of the administration of tests; and (3) to train them in the interpretation and utilization of a few standard tests. Emphasis during the past biennium has, perhaps been given to the third objective."

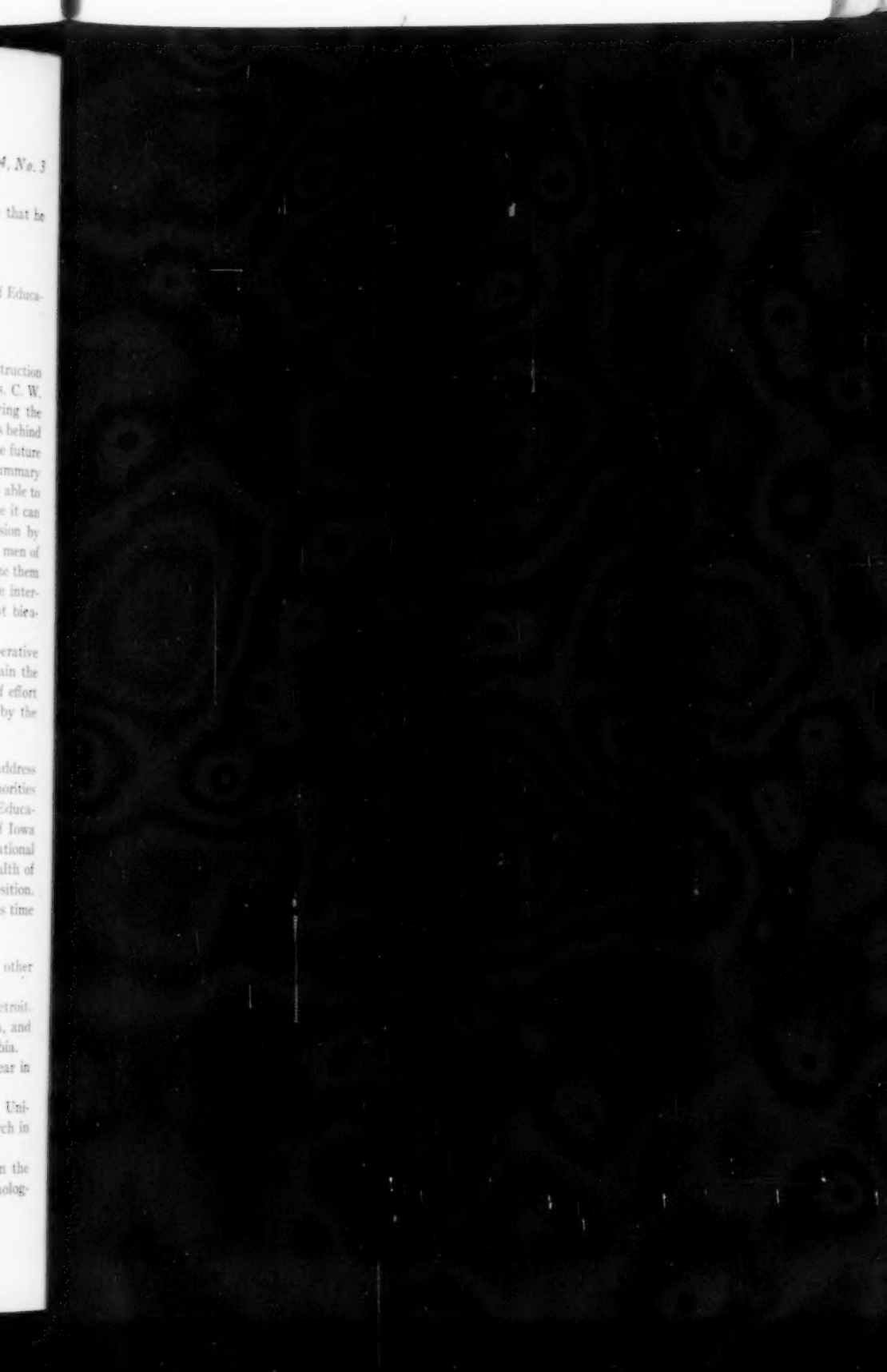
Visitation, conferences, lectures, demonstrations, institutes, state-wide cooperative studies and preparation of special aids and reports have all been used to attain the desired ends. From the report it would seem that the tremendous amount of effort expended has been justified in part by the results accomplished and in part by the foundation laid for future activity.

Perhaps this is the proper place for the announcement of the change of address of your secretary. On invitation of Dr. Buckingham and the other proper authorities of Ohio State University to become Assistant Director of the new Bureau of Educational Research, Dr. Ashbaugh resigned his position at the State University of Iowa to accept this invitation. Dr. Ashbaugh was in charge of the Bureau of Educational Service in the Extension Division there for seven years where he gained a wealth of experience which will be directly applicable to the problems in his new position. During the first year of organization of the new bureau he will give a part of his time to teaching "School Administration" in the College of Education.

Announcement should also be made of change of position or address of other members.

Mr. P. C. Packer has resigned as assistant superintendent of schools in Detroit. He is teaching this first semester at his Alma Mater, State University of Iowa, and will spend the second semester in graduate study at Teachers College, Columbia.

Mr. W. W. Coxe of the Vocational Bureau, Cincinnati is spending this year in graduate study at Ohio State University.

Dr. Chas. Fordyce has resigned the deanship of Teachers College at the University of Nebraska to devote his entire time to the work of educational research in that institution.

Miss Henrietta V. Race, formerly Director of Intelligence Measurement in the Schools of Kansas City, Missouri, has become the Director of the Bureau of Psychological and Educational Measurement at Youngstown, Ohio.